# CANCER ASSOCIATED ANTIGENS, SGA-56M AND SGA-56Mv, ~~AND USES THEREOF~~

[0001] This application claims priority under 35 USC §119(e) from U.S Provisional Application Serial No. 60/410,048 filed 12 September 2002, which application is herein specifically incorporated by reference in its entirety.

## 1. FIELD OF THE INVENTION

[0002] The invention relates generally to the field of cancer diagnosis, prognosis, treatment and prevention. More particularly, the present invention relates to methods of diagnosing, treating and preventing breast cancer. Methods of using a nucleic acid and a protein, differentially expressed in tumor cells, and antibodies against the protein, to treat, diagnose or prevent cancer, are provided for by the present invention. The instant invention provides compositions comprising, and methods of using, products of a gene termed SGA-56M and variants thereof, including SGA-56Mv. Such SGA-56M gene products include SGA-56M proteins and nucleic acids and variants thereof, including SGA-56Mv. Such gene products, as well as their binding partners and antagonists, can be used for the prevention, diagnosis, prognosis and treatment of cancer.

## 2. BACKGROUND OF THE INVENTION

[0003] Cancer is characterized primarily by an increase in the number of abnormal cells derived from a given normal tissue, invasion of adjacent tissues by these abnormal cells, and lymphatic or blood-borne spread of malignant cells to regional lymph nodes and to distant sites (metastases). Clinical data and molecular biologic studies indicate that cancer is a multistep process that begins with minor preneoplastic changes, which may under certain conditions progress to neoplasia.

[0004] Pre-malignant abnormal cell growth is exemplified by hyperplasia, metaplasia, or most particularly, dysplasia (for review of such abnormal growth conditions, *see* Robbins & Angell, 1976, *Basic Pathology*, 2d Ed., W.B. Saunders Co., Philadelphia, pp. 68-79) The neoplastic lesion may evolve clonally and develop an increasing capacity for growth, metastasis, and heterogeneity, especially under conditions in which the neoplastic cells escape the host's immune surveillance (Roitt, I., Brostoff, J. and Kale, D., 1993, *Immunology*, 3rd ed., Mosby, St. Louis, pps. 17.1-17.12).

[0005] The incidence of breast cancer, a leading cause of death in women, has been gradually increasing in the United States over the last thirty years. Its cumulative risk is relatively high, 1 in 8 women, for example, by age 85 in the United States. In fact, breast cancer is the most common cancer in women and the second most common cause of cancer death in the United States. In 1997, it was estimated that 181,000 new cases were reported in the U.S., and that 44,000 people would die of breast cancer (Parker *et al.*, 1997, *CA Cancer J. Clin.* 47:5; Chu *et al.*, 1996, *J. Nat. Cancer Inst.* 88:1571). While the mechanism of tumorigenesis for most breast carcinomas is largely unknown, there are genetic factors that can predispose some women to developing breast cancer (Miki *et al.*, 1994, *Science* 266:66). The discovery and characterization of *BRCA1* and *BRCA2* has expanded our knowledge of genetic factors that can contribute to familial breast cancer.

Germ-line mutations within these two loci are associated with a 50 to 85% lifetime risk of breast and/or ovarian cancer (Casey, 1997, *Curr. Opin. Oncol.* 9: 88; Marcus *et al.*, 1996, *Cancer* 77: 697). Sporadic tumors, however, or those not known to be associated with a known germline mutation, constitute the majority of breast cancers. It is likely that other, non-genetic factors also have a significant effect on the etiology of the disease. Regardless of its origin, breast cancer morbidity and mortality increases significantly if it is not detected early in its progression. Thus, considerable effort has focused on the early detection of cellular transformation and tumor formation in breast tissue.

**[0006]** Only about 5% to 10% of breast cancers are associated with breast cancer susceptibility genes, *BRCA1* and *BRCA2*. The cumulative lifetime risk of breast cancer for women who carry the mutant *BRCA1* is predicted to be approximately 92%, while the cumulative lifetime risk for the non-carrier majority is estimated to be approximately 10%. *BRCA1* is a tumor suppressor gene that is involved in DNA repair and cell cycle control, which are both important for the maintenance of genomic stability. More than 90% of all mutations reported so far result in a premature truncation of the protein product with abnormal or abolished function. The histology of breast cancer in *BRCA1* mutation carriers differs from that in sporadic cases, but mutation analysis is the only way to identify a carrier. Like *BRCA1*, *BRCA2* is involved in the development of breast cancer and plays a role in DNA repair. However, unlike *BRCA1*, it is not involved in ovarian cancer.

**[0007]** Other genes have been linked to breast cancer, for example c-erb-2 (*HER2*) and p53 (Beenken *et al.* 2001, *Ann. Surg.* 233(5):630). Over expression of c-erb-2 (*HER2*) and p53 have been correlated with poor prognosis (Rudolph *et al.* 2001, *Hum. Pathol.* 32(3):311), as has been aberrant expression products of *mdm2* (Lukas *et al. 2001*, *Cancer Res.* 61(7):3212) and cyclin1 and p27 (Porter & Roberts, International Publication WO98/33450, published August 6, 1998).

**[0008]** A marker-based approach to tumor identification and characterization promises improved diagnostic and prognostic reliability. Typically, the diagnosis of breast cancer and other types of cancer requires histopathological proof of the presence of the tumor. In addition to diagnosis, histopathological examinations also provide information about prognosis and selection of treatment regimens. Prognosis may also be established based upon clinical parameters such as tumor size, tumor grade, the age of the patient, and lymph node metastasis.

**[0009]** In clinical practice, accurate diagnosis of various subtypes of cancer is important because treatment options, prognosis, and the likelihood of therapeutic response all vary broadly depending on the diagnosis. Accurate prognosis, or determination of distant metastasis-free survival could allow the oncologist to tailor the administration of adjuvant chemotherapy, with patients having poorer prognoses being given the most aggressive treatment. Furthermore, accurate prediction of poor prognosis would greatly impact clinical trials for new breast cancer therapies, because potential study patients could then be stratified according to prognosis. Trials could then be limited to patients having poor prognosis, in turn making it easier to discern if an experimental therapy is efficacious. To date, no set of satisfactory predictors for prognosis based on the clinical information alone has been identified. The detection of *BRCA1* or *BRCA2* mutations represents a step towards the design of improved therapeutics and therapeutic regimens for preventing and regulating the appearance of these tumors.

**[0010]** It would, therefore, be beneficial to provide specific methods and reagents for the diagnosis, staging, prognosis, monitoring and treatment of cancer, including breast cancer, and to provide methods for identifying individuals with a predisposition for the onset of breast cancer, and other types of cancer, and hence are appropriate subjects for preventive therapy.

## 3. SUMMARY OF THE INVENTION

**[0011]** Intensive and systematic evaluation of gene expression patterns is essential in understanding the physiological mechanisms associated with cellular transformation and metastasis associated with cancer. Several techniques that permit comparison of gene expression in normal and cancerous cells are known in the art. Examples of these techniques include: Serial Analysis of Gene Expression (SAGE) (Velculescu *et al.*, 1995, *Science* 270:484); Restriction Enzyme Analysis of Differentially Expressed Sequences (READS) (Prasher *et al.*, 1999, *Methods in Enzymology* 303:258); Amplified Fragment Length Polymorphism (AFLP) (Bachem *et al.*, 1996, *Plant Journal* 9:745); Representational Difference Analysis (RDA) (Hubank *et al.*, 1994, *Nucleic Acid Research* 22:(25):5640); differential display (Liang *et al.*, 1992, *Cancer Research* 52(24):6966); and suppression subtractive hybridization (SSH) (Diatchenko *et al.*, 1996, *Proc. Natl. Acad. Sci. USA* 93:6025). As described herein, the present inventors have used differential expression methods to identify and characterize the SGA-56M gene and variants thereof, including SGA-56Mv, as genes whose expression is associated with breast cancer and other types of cancer. This discovery by the present inventors has made possible the use of SGA-56M and variants thereof, including SGA-56Mv, for the treatment, prevention and diagnosis of cancers, including but not limited to breast cancer.

**[0012]** The present invention relates to the discovery that a gene, SGA-56M and variants thereof, including SGA-56Mv, has an expression pattern that is up-regulated in cancer tissues and cell lines, *e.g.*, breast cancer tissues and cell lines. The invention relates to the use of said gene, gene products, and antagonists of said gene or gene products (SGA-56M and variants thereof, including SGA-56Mv, cDNA, RNA, and/or protein) as targets for diagnosis, drug screening and therapies for cancer. The present invention also relates to the use of said genes or gene products or derivatives thereof as vaccines against cancer. In a preferred embodiment, the invention provides for methods of using the protein, SGA-56M and variants thereof, including SGA-56Mv, or nucleic acids that encode said proteins for the treatment, prevention and diagnosis of breast cancer.

**[0013]** In particular, the methods of the present invention include using nucleic acid molecules that encode the SGA-56M protein and variants thereof, including SGA-56Mv, and recombinant DNA molecules, cloned genes or degenerate variants thereof, and in particular naturally occurring variants that encode SGA-56M related gene products. The methods of the present invention additionally include using cloning vectors, including expression vectors, comprising the nucleic acid molecules encoding SGA-56M and variants thereof (e.g., SGA-56Mv), and hosts that comprise such nucleic acid molecules. The methods of the present invention also encompass the use of SGA-56M gene products and variants thereof, including SGA-56Mv, fusion proteins, and antibodies directed against such SGA-56M gene products or conserved variants or fragments thereof. In one embodiment, a fragment or other derivative of an SGA-56M protein is at least 10

amino acids long. In another embodiment, a fragment of an SGA-56M nucleic acid and variants thereof, including SGA-56Mv, nucleic acid or derivative thereof is at least 10 nucleotides long.

**[0014]** The nucleotide sequence of the cDNA of a human SGA-56M gene, and SGA-56Mv is provided. The nucleotide sequences of the SGA-56M ORF, and SGA-56Mv ORF in the SGA-56M, and SGA-56Mv genes, as well as the amino acid sequences of the encoded gene products, are also provided. The SGA-56M and SGA-56Mv genes were cloned by PCR. The SGA-56M transcript encodes a protein of 802 amino acids and the SGA-56Mv transcript encodes a protein of 756 amino acids. The SGA-56Mv protein has an in-frame deletion of amino acids 234-280 of SGA-56M. In-frame start and stop sequences were observed by sequence analysis of the SGA-56M and SGA-56Mv genes. The SGA-56M and SGA-56Mv transcripts were both detected at elevated levels in both breast cancer cell-lines and breast tumor isolates as compared to normal tissues. Elevated transcript levels of SGA-56M and SGA-56Mv genes were also associated with lung cancer tissue. Elevated transcript levels were also detected in several other tumor types and cancer cells as described in Section 6.

**[0015]** The present invention further relates to methods for the diagnostic evaluation and prognosis of cancer in a subject animal. Preferably the subject is a mammal, more preferably the subject is a human. In a preferred embodiment the invention relates to methods for diagnostic evaluation and prognosis of breast cancer. For example, nucleic acid molecules of the invention can be used as diagnostic hybridization probes or as primers for diagnostic PCR analysis for detection of abnormal expression of SGA-56M and SGA-56Mv genes.

**[0016]** Antibodies or other specific binding partners to the SGA-56M and variants thereof, including SGA-56Mv proteins, of the invention can be used in a diagnostic test to detect the presence of the SGA-56M or SGA-56Mv gene products in body fluids, cells or in tissue biopsy. In specific embodiments, measurement of serum or cellular SGA-56M and variants thereof, including SGA-56Mv protein levels can be made to detect or stage breast cancer, *e.g.*, infiltrative ductal carcinoma.

**[0017]** The present invention also relates to methods for the identification of subjects having a predisposition to cancer, *e.g.*, breast cancer. The subject can be any animal, but preferably the subject is a mammal, and most preferably the subject is a human. In a non-limiting example nucleic acid molecules of the invention can be used as diagnostic hybridization probes or as primers for quantitative RT- PCR analysis to determine expression levels of the SGA-56M gene products and variants thereof, including SGA-56Mv. In another example, nucleic acid molecules of the invention can be used as diagnostic hybridization probes or as primers for diagnostic PCR analysis for the identification of SGA-56M and variants thereof, including SGA-56Mv naturally occurring or non-naturally occurring gene mutations, allelic variations and regulatory defects in SGA-56M and SGA-56Mv genes.

**[0018]** Imaging methods, for imaging the localization and/or amounts of SGA-56M and variants thereof, including SGA-56Mv gene products in a patient, are also provided for diagnostic and prognostic use.

**[0019]** Further, methods are presented for the treatment of cancer, including breast cancer. Such methods comprise the administration of compositions that are capable of modulating the level of SGA-56M and

variants thereof, including SGA-56Mv gene expression and/or the level of SGA-56M and SGA-56Mv gene product activity in a subject. The subject can be any animal, preferably a mammal, more preferably a human.

**[0020]** Still further, the present invention relates to methods for the use of the SGA-56M gene and variants thereof, including SGA-56Mv, and/or SGA-56M and variants thereof, including SGA-56Mv gene products for the identification of compounds that modulate SGA-56M or SGA-56Mv gene expression and/or the activity of SGA-56M or SGA-56Mv gene products. Such compounds can be used as agents to prevent and/or treat breast cancer or any cancer wherein SGA-56M and variants thereof, including SGA-56Mv, are expressed at levels that are higher than those detected in corresponding normal tissue. Such compounds can also be used to palliate the symptoms of the disease, and control the metastatic potential of breast cancer or any cancer wherein SGA-56M and variants thereof, including SGA-56Mv, are expressed at levels higher than those observed in corresponding normal tissue.

**[0021]** The invention also provides methods of preventing cancer by administering the product of the SGA-56M gene and variants thereof, including SGA-56Mv or a fragment of the SGA-56M gene product and variants thereof, including SGA-56Mv in an amount effective to elicit an immune response in a subject. The subject can be any animal, preferably a mammal, more preferably a human. The invention also provides methods of treating or preventing cancer by administering the nucleic acid that encodes the SGA-56M gene product and variants thereof, including SGA-56Mv or a fragment of the nucleic acid that encodes the SGA-56M or SGA-56Mv gene product in an amount effective to elicit an immune response. The invention further provides methods of treating or preventing cancer by administering a protein or a peptide encoded by the SGA-56M gene and variants thereof, including SGA-56Mv in an amount effective to elicit an immune response. The immune response can be either humoral or cellular or both. In a preferred embodiment the invention provides a method of immunizing against breast or lung cancer.

**[0022]** The invention relates to screening assays to identify antagonists or agonists of the SGA-56M gene or gene product and variants thereof, including SGA-56Mv. Thus, the invention relates to methods of identifying agonists or antagonists of the SGA-56M gene or gene product and variants thereof, including SGA-56Mv and the use of said agonist or antagonist to treat or prevent breast cancer or other types of cancer.

**[0023]** The invention also provides methods of treating cancer by providing therapeutic amounts of an anti-sense nucleic acid molecule. An anti-sense nucleic molecule is a nucleic acid molecule that is the complement of all or a part of the SGA-56M or SGA-56Mv gene sequences or SGA-56M and SGA-56Mv ORFs and which therefore can hybridize to the SGA-56M gene and variants thereof, including SGA-56Mv or a fragment thereof. Hybridization of the anti-sense molecule can inhibit expression of the SGA-56M or SGA-56Mv gene. In a preferred embodiment the method is used to treat breast cancer.

**[0024]** The invention also includes a kit for assessing whether a patient is afflicted with breast cancer or other types of cancer. This kit comprises reagents for assessing expression of an SGA-56M or SGA-56Mv gene product.

**[0025]** In another aspect, the invention relates to a kit for assessing the suitability of each of a plurality of compounds for inhibiting cancer including breast cancer in a patient. The kit comprises a reagent for

assessing expression of an SGA-56M or SGA-56Mv gene products, and may also comprise a plurality of compounds.

**[0026]** In another aspect, the invention relates to a kit for assessing the presence of cancer cells. This kit comprises an antibody, wherein the antibody binds specifically with a protein corresponding to an SGA-56M gene product and variants thereof, including SGA-56Mv. The kit may also comprise a plurality of antibodies, wherein the plurality binds specifically with different epitopes on an SGA-56M gene product and variants thereof, including SGA-56Mv.

**[0027]** The invention also includes a kit for assessing the presence of cancer cells, wherein the kit comprises a nucleic acid (*e.g.*, oligonucleotide) probe. The probe binds specifically with a transcribed polynucleotide corresponding to an SGA-56M gene product and variants thereof, including SGA-56Mv. The kit may also comprise a plurality of probes, wherein each of the probes binds specifically with a transcribed polynucleotide corresponding to a different mRNA sequence transcribed from the SGA-56M gene and variants thereof, including SGA-56Mv.

**[0028]** Kits for diagnostic use, comprising in a container, primers for use in PCR that can amplify SGA-56M cDNA and variants thereof, including the SGA-56Mv cDNA and/or genes and, in a separate container, a standard amount of SGA-56M or SGA-56Mv cDNA are also provided.

**[0029]** The invention also provides transgenic non-human animals (*e.g.*, mice) that express SGA-56M or SGA-56Mv nucleic acids and proteins encoded by a transgene. Transgenic, non-human knockout animals (*e.g.*, mice), in which an SGA-56M gene and variants thereof, including SGA-56Mv has been partially or completely inactivated, are also provided.

**[0030]** Accordingly, the present invention provides a method for diagnosing a cancer in a subject comprising detecting or measuring an SGA-56M or SGA-56Mv gene product in a sample derived from said subject, wherein said SGA-56M or SGA-56Mv gene product is (a) an RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4; (b) a protein comprising SEQ ID NO: 5 or SEQ ID NO: 6; (c) a nucleic acid comprising a sequence hybridizable to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, under conditions of high stringency, or a protein comprising a sequence encoded by said hybridizable sequence; (d) a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, as determined using the NBLAST algorithm, or a protein encoded thereby; in which elevated expression levels of an SGA-56M gene product and/or variants thereof, including SGA-56Mv, relative to those of a non-cancerous sample or a pre-determined standard value for a noncancerous sample, indicate the presence of cancer in the subject. In one embodiment of the foregoing diagnostic method, the subject is a human. In another embodiment, the cancer is breast or lung cancer. In yet other embodiments, the sample is a tissue sample, a plurality of cells, or a bodily fluid.

**[0031]** The present invention further provides methods for staging a cancer in a subject comprising detecting or measuring an SGA-56M gene product and/or variants thereof, including SGA-56Mv, in a sample derived from said subject, wherein said SGA-56M gene product and/or variants thereof, (e.g., SGA-56Mv) is (a) an RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4; (b) a protein

comprising SEQ ID NO: 5 or SEQ ID NO: 6; (c) a nucleic acid comprising a sequence hybridizable to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, under conditions of high stringency, or a protein comprising a sequence encoded by said hybridizable sequence; (d) a nucleic acid comprising a sequence hybridizable to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, under conditions of high stringency, or a protein comprising a sequence encoded by said hybridizable sequence; (e) a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, as determined using the NBLAST algorithm, or a protein encoded thereby; in which elevated expression levels of the SGA-56M gene product and/or variants thereof, including SGA-56Mv, relative to those of a non-cancerous sample or a pre-determined standard value for a noncancerous sample, indicate an advanced stage of cancer in the subject.

**[0032]** The present invention further provides methods for the treatment of a cancer in a subject, comprising administering to the subject a therapeutically effective amount of a compound for treating the cancer that antagonizes an SGA-56M gene product and/or variants thereof, including SGA-56Mv, wherein said SGA-56M or SGA-56Mv gene product is (a) an RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4; (b) a protein comprising SEQ ID NO: 5 or SEQ ID NO: 6; (c) a nucleic acid comprising a sequence hybridizable to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, under conditions of high stringency, or a protein comprising a sequence encoded by said hybridizable sequence; or a protein comprising a sequence encoded by said hybridizable sequence; (d) a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, as determined using the NBLAST algorithm; or a protein encoded thereby. In one embodiment, the gene product whose expression is down-regulated is a protein encoded by a nucleic acid comprising a nucleotide sequence with at least 90% sequence identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4. In another embodiment, the compound decreases expression of an RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4. The antagonist can be (i) a protein; (ii) a peptide; (iii) an organic molecule with a molecular weight of less than 500 daltons; (iv) an inorganic molecule with a molecular weight of less than 500 daltons; (v) an antisense oligonucleotide molecule that binds to said RNA and inhibits translation of said RNA; (vi) a ribozyme molecule that targets said RNA and inhibits translation of said RNA; (vii) an antibody that specifically/selectively binds to an SGA-56M gene product and variants thereof, including SGA-56Mv; or (viii) a double stranded oligonucleotide that forms a triple helix with a promoter of an SGA-56M gene and variants thereof, including SGA-56Mv, wherein said SGA-56M and SGA-56Mv gene is a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, as determined using the NBLAST algorithm. In an embodiment wherein the compound is an antibody, the antibody immunospecifically binds to a protein comprising the amino acid sequence of SEQ ID NO: 5 or SEQ ID NO: 6.

**[0033]** The present invention further provides methods for vaccinating a subject against cancer comprising administering to the subject a molecule that elicits an immune response to an SGA-56M and/or SGA-56Mv

gene product, wherein said SGA-56M and/or SGA-56Mv gene product is (a) an RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4; (b) a protein comprising SEQ ID NO: 5 or SEQ ID NO: 6 (c) a nucleic acid comprising a sequence hybridizable to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, under conditions of high stringency, or a protein comprising a sequence encoded by said hybridizable sequence; (d) a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, as determined using the NBLAST algorithm, or a protein encoded thereby. In one embodiment, the immune response is a cellular immune response. In another embodiment, the immune response is a humoral immune response. In yet another embodiment, the immune response is both a cellular and a humoral immune response. Such immune responses confer protective immunity against a cancer to a patient. Protective immunity refers to a reduced risk for developing a cancer and, therefore, encompasses a partial or complete immunity.

**[0034]** The present invention yet further provides methods for determining risk of developing cancer in a subject, said method comprising (I) measuring an amount of an SGA-56M and/or SGA-56Mv gene product in a sample derived from the subject, wherein said SGA-56M and/or SGA-56Mv gene product is: (a) an RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4; (b) a protein comprising SEQ ID NO: 5 or SEQ ID NO: 6; (c) a nucleic acid comprising a sequence hybridizable to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, under conditions of high stringency, or a protein comprising a sequence encoded by said hybridizable sequence; (d) a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, as determined using the NBLAST algorithm, or a protein encoded thereby; and (II) comparing the amount of said SGA-56M and/or SGA-56Mv gene product in the subject with the amount of SGA-56M and/or SGA-56Mv gene product present in a non-cancerous sample or predetermined standard for a noncancerous sample, wherein an elevated amount of said SGA-56M or SGA-56Mv gene product in the subject relative to the amount in the non-cancerous sample or pre-determined standard for a noncancerous sample indicates a risk of developing cancer in the subject.

**[0035]** The present invention yet further provides methods for determining if a subject suffering from cancer is at risk for metastasis of said cancer, said method comprising measuring an amount of an SGA-56M and/or SGA-56Mv gene product in a sample derived from the subject, wherein said gene product is (a) an RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4; (b) a protein comprising SEQ ID NO: 5 or SEQ ID NO: 6; (c) a nucleic acid comprising a sequence hybridizable to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, under conditions of high stringency, or a protein comprising a sequence encoded by said hybridizable sequence; (d) a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, as determined using the NBLAST algorithm, or a protein encoded thereby, wherein an elevated amount of SGA-56M or SGA-56Mv gene products in the subject compared to the amount in the non-cancerous sample, or in the sample from the subject with the non-metastasizing cancer, or the amount in

the predetermined standard for a noncancerous or non-metastasizing sample, indicates an increased risk for developing metastasis of said cancer in the subject.

**[0036]** The present invention yet further provides methods for screening to identify a compound capable of binding to an SGA-56M or SGA-56Mv molecule, said method comprising (I) contacting the SGA-56M or SGA-56Mv molecule with a candidate agent, wherein said SGA-56M or SGA-56Mv molecule is (a) an RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4; (b) a protein comprising SEQ ID NO: 5 or SEQ ID NO: 6; (c) a nucleic acid comprising a sequence hybridizable to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, under conditions of high stringency, or a protein comprising a sequence encoded by said hybridizable sequence; (d) a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, as determined using the NBLAST algorithm, or a protein encoded thereby and (II) determining if the candidate agent binds the SGA-56M or SGA-56Mv molecule. The screening assay can be performed *in vitro*. In one embodiment, the SGA-56M molecule, and/or variants thereof (e.g., SGA-56Mv) is anchored to a solid phase. In another embodiment, the candidate agent is anchored to a solid phase. In other embodiments, the screening assay is performed in solution. In yet other embodiments, the SGA-56M molecule or variant thereof (e.g., SGA-56Mv) is expressed on the surface of a cell or in the cytosol of a cell in step (I). In other embodiments, the SGA-56M molecule or variant thereof (e.g., SGA-56Mv) is expressed endogenously in the cell; alternatively, the cell can be engineered to express exogenous SGA-56M and/or variants thereof. In the foregoing screening methods, the candidate agent is preferably labeled, for example radioactively or enzymatically.

**[0037]** The present invention also encompasses a method for screening to identify a protein capable of interacting with an SGA-56M gene product comprising contacting an SGA-56M gene product to a plurality of polypeptides, wherein the SGA-56M gene product is: an RNA corresponding to SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, or SEQ ID NO:4; a protein comprising SEQ ID NO:5 or SEQ ID NO:6; a nucleic acid comprising a sequence hybridizable to SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, or SEQ ID NO:4, or a complement thereof, under conditions of high stringency, or a protein comprising a sequence encoded by said hybridizable sequence; a nucleic acid at least 90% homologous to SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, or SEQ ID NO:4, or a complement thereof, as determined using an NBLAST algorithm, or a protein encoded thereby; or a nucleic acid sequence encoding a protein comprising SEQ ID NO:5 or SEQ ID NO:6; and determining if at least one protein binds to or forms a complex with the SGA-56M gene product. Such methods may be performed *in vitro* or *in vivo*, for example, in a cell. In one embodiment, the method for screening is a two-hybrid screening method which is generally performed in yeast cells.

**[0038]** The present invention provides methods for screening to identify a protein or peptide that interacts with an SGA-56M or SGA-56Mv gene product, said method comprising (I) immunoprecipitating the SGA-56M or SGA-56Mv gene product from a cell lysate, wherein said SGA-56M or SGA-56Mv gene product is (a) an RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4; (b) a protein comprising SEQ ID NO: 5 or SEQ ID NO: 6; (c) a nucleic acid comprising a sequence hybridizable

to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, under conditions of high stringency, or a protein comprising a sequence encoded by said hybridizable sequence; (d) a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, as determined using the NBLAST algorithm, or a protein encoded thereby; and (II) determining if at least one cellular protein binds to or forms a complex with the SGA-56M or SGA-56Mv gene product in the immunoprecipitate.

**[0039]** The present invention yet further provides methods for screening to identify a candidate agent capable of modulating (e.g., decreasing or increasing) the expression level and/or activity of an SGA-56M gene, and/or variant thereof, such as SGA-56Mv, said method comprising (I) contacting said SGA-56M or SGA-56Mv gene with a candidate agent, wherein said SGA-56M or SGA-56Mv gene is a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4 as determined using the NBLAST algorithm; and (II) measuring the level of expression and/or activity of an SGA-56M or SGA-56Mv gene product, said SGA-56M or SGA-56Mv gene product selected from the group consisting of an mRNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4 or a protein comprising SEQ ID NO: 5 or SEQ ID NO: 6, wherein an increase or decrease in said level of expression and/or activity relative to said level of expression and/or activity in the absence of said candidate agent indicates that the candidate agent modulates expression of an SGA-56M or SGA-56Mv gene.

**[0040]** The present invention yet further provides an immunogenic composition comprising (I) a purified SGA-56M or SGA-56Mv gene product in an amount effective at eliciting an immune response, wherein said gene product is (a) an RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4; (b) a protein comprising SEQ ID NO: 5 or SEQ ID NO: 6; (c) a nucleic acid comprising a sequence hybridizable to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, under conditions of high stringency, or a protein comprising a sequence encoded by said hybridizable sequence; (d) a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, as determined using the NBLAST algorithm, or a protein encoded thereby; and (II) a pharmaceutically acceptable carrier or excipient.

**[0041]** The present invention yet further provides a pharmaceutical composition comprising an antibody that specifically binds to a protein comprising SEQ ID NO:5 or SEQ ID NO:6; and a pharmaceutically acceptable carrier.

**[0042]** The present invention yet further provides pharmaceutical compositions comprising (I) an SGA-56M or SGA-56Mv gene product (e.g., SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 or SEQ ID NO:4) or an antibody immunospecific for a protein comprising SEQ ID NO:5 or SEQ ID NO:6; and (II) a pharmaceutically acceptable carrier.

**[0043]** The pharmaceutical compositions of the present invention can be formulated, *inter alia*, for delivery as an aerosol, for parenteral delivery, or for oral delivery.

**[0044]** The present invention yet further provides methods for diagnosing cancer in a subject comprising (I) administering to said subject a compound that binds specifically to a protein comprising the amino acid

sequence of SEQ ID NO: 5 or SEQ ID NO: 6, wherein said compound is bound to an imaging agent; and (II) obtaining an internal image of said subject by visualizing the compound bound to the imaging agent, wherein the detection of the compound bound to the imaging agent provides a positive indicator for diagnosing a cancer in the subject. In a preferred embodiment, the compound is an antibody. In a preferred mode of the embodiment, the antibody is conjugated to a radioactive metal and said obtaining step comprises recording a scintographic image obtained from the decay of the radioactive metal.

**[0045]** Also provided are kits that are useful for practicing the methods of the present invention. In one embodiment, such a kit comprises, in one or more containers, an oligonucleotide primer pair, wherein each primer is complementary to a different strand of a double-stranded nucleic acid sequence comprising SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, wherein said primer pair is capable of priming a DNA amplification reaction; and, in a separate container, a reference DNA comprising a purified double-stranded nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4. In specific modes of the embodiment, each primer comprises a nucleotide sequence with at least 8, more preferably at least 10, yet more preferably at least 12, and most preferably at least 15 complementary nucleotides to complementary strands of a double-stranded nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4.

**[0046]** The present invention yet further provides a transgenic non-human animal which expresses from a transgene an SGA-56M or SGA-56Mv gene product, for example, an RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or a protein comprising SEQ ID NO: 5 or SEQ ID NO: 6.

**[0047]** The present invention yet further provides a method for testing the effects of a candidate therapeutic compound comprising administering said compound to a transgenic non-human animal which expresses an exogenous SGA-56M or SGA-56Mv gene product; and determining any effects of said compound upon said transgenic non-human animal.

**[0048]** The present invention further provides host cells comprising nucleic acids encoding the polypeptides of the invention operably linked to a promoter, and methods of expressing such polypeptides by culturing the host cells under conditions in which the nucleic acid molecule is expressed.

## 3.1 DEFINITIONS

**[0049]** SPECIFIC OR SELECTIVE: a nucleic acid used in a reaction, such as a probe used in a hybridization reaction, a primer used in a PCR, or a nucleic acid present in a pharmaceutical preparation, is referred to as "selective" if it hybridizes or reacts with the intended target more frequently, more rapidly, or with greater duration than it does with alternative substances. Similarly, a polypeptide is referred to as "selective" if it binds an intended target, such as a ligand, hapten, substrate, antibody, or other polypeptide more frequently, more rapidly, or with greater duration than it does to alternative substances. An antibody is referred to as "selective" if it binds via at least one antigen recognition site to the intended target more frequently, more rapidly, or with greater duration than it does to alternative substances. A marker is selective to a particular

cell or tissue type if it is expressed predominantly in or on that cell or tissue type, particularly with respect to a biological sample of interest.

**[0050]** VARIANT (S): A variant (v) of a polynucleotide or polypeptide, as the term is used herein, is a polynucleotide or polypeptide that is different from a reference polynucleotide or polypeptide, respectively.

**[0051]** Variant polynucleotides are generally limited so that the nucleotide sequence of the reference and the variant are closely related overall and, in many regions, identical. Changes in the nucleotide sequence of the variant may be silent. That is, they may not alter the amino acid sequence encoded by the polynucleotide. Where alterations are limited to silent changes of this type a variant will encode a polypeptide with the same amino acid sequence as the reference. Alternatively, changes in the nucleotide sequence of the variant may alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Such nucleotide changes may result in amino acid substitutions, additions, deletions, fusions, and truncations in the polypeptide encoded by the reference sequence.

**[0052]** Variant polypeptides are generally limited so that the sequences of the reference and the variant are similar overall and, in many regions, identical. For example, a variant and reference polypeptide may differ in amino acid sequence by one or more substitutions (conservative or non-conservative), additions, deletions, fusions, and truncations, which may be present or absent in any combination.

**[0053]** CORRESPOND OR CORRESPONDING: Between nucleic acids, "corresponding" means homologous to or complementary to a particular sequence or portion of the sequence of a nucleic acid. As between nucleic acids and polypeptides, "corresponding" refers to amino acids of a peptide encoded by the nucleic acid sequence or a portion thereof or a complement of either. As between polypeptides (e.g., peptides, polypeptides, or proteins), "corresponding" refers to an amino acid sequence of a first polypeptide that is identical or homologous to an amino acid sequence of a second polypeptide.

**[0054]** SGA-56M GENE PRODUCT: As used herein, unless otherwise indicated, an SGA-56M gene product is: an RNA corresponding to SEQ ID NO: 1 or SEQ ID NO: 2; a protein comprising SEQ ID NO: 5; a nucleic acid sequence encoding an amino acid sequence comprising SEQ ID NO: 5; a nucleic acid comprising a sequence hybridizable to SEQ ID NO: 1 or SEQ ID NO: 2, or complement thereof, under conditions of high stringency, or a protein comprising a sequence encoded by said hybridizable sequence; a nucleic acid at least 90% homologous to SEQ ID NO: 1 or SEQ ID NO: 2, or a complement thereof, as determined using the NBLAST algorithm; a nucleic acid at least 90% homologous to SEQ ID NO: 1 or SEQ ID NO: 2, or a complement thereof, or a fragment or derivative of any of the foregoing proteins or nucleic acids.

**[0055]** SGA-56Mv GENE PRODUCT: As used herein, unless otherwise indicated, an SGA-56Mv gene product is: an RNA corresponding to SEQ ID NO: 3 or SEQ ID NO: 4; a protein comprising SEQ ID NO: 6; a nucleic acid sequence encoding an amino acid sequence comprising SEQ ID NO: 6; a nucleic acid comprising a sequence hybridizable to SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, under conditions of high stringency, or a protein comprising a sequence encoded by said hybridizable sequence; a nucleic acid at least 90% homologous to SEQ ID NO: 3 or SEQ ID NO: 4, or a complement thereof, as determined using the

NBLAST algorithm; a nucleic acid at least 90% homologous to SEQ ID NO: 3 or SEQ ID NO: 4 or a fragment or derivative of any of the foregoing proteins or nucleic acids.

**[0056]** CONTROL ELEMENTS: As used herein refers collectively to promoter regions, polyadenylation signals, transcription termination sequences, upstream regulatory domains, origins of replication, internal ribosome entry sites ("IRES"), enhancers, and the like, which collectively provide for the replication, transcription and translation of a coding sequence in a recipient cell. Not all of these control elements need always be present so long as the selected coding sequence is capable of being replicated, transcribed and translated in an appropriate host cell.

**[0057]** PROMOTER REGION: Is used herein in its ordinary sense to refer to a nucleotide region comprising a DNA regulatory sequence, wherein the regulatory sequence is derived from a gene which is capable of binding RNA polymerase and initiating transcription of a downstream (3'-direction) coding sequence.

**[0058]** OPERABLY LINKED: As used herein refers to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, control elements operably linked to a coding sequence are capable of effecting the expression of the coding sequence. The control elements need not be contiguous with the coding sequence, so long as they function to direct the expression thereof.

**[0059]** MODULATE: As used herein, a compound which is capable of increasing or decreasing the level and/or activity of an SGA-56M and/or SGA-56Mv molecule may be referred to herein as an SGA-56M and/or SGA-56Mv modulator.

**[0060]** ANTAGONIST: As used herein, a compound capable of reducing the level and/or activity of an SGA-56M and/or SGA-56Mv molecule or a variant thereof may be referred to herein as an SGA-56M and/or SGA-56Mv antagonist.

**[0061]** AGONIST: As used herein, a compound capable of increasing the level and/or activity of an SGA-56M and/or SGA-56Mv molecule or a variant thereof may be referred to herein as an SGA-56M and/or SGA-56Mv agonist.

**[0062]** SGA-56M AND/OR SGA-56Mv ACTIVITY: As used herein, the term "SGA-56M and/or SGA-56Mv activity" refers to an activity of SGA-56M and/or SGA-56Mv which contributes to the onset, progression, and/or metastatic spread of breast or lung cancer and/or other cancers.

**[0063]** ELEVATED SGA-56M and/or SGA-56Mv LEVELS: As used herein the terms "elevated", "over-expressed", "up-regulated", or "increased" SGA-56M and/or SGA-56Mv levels refer to an approximately two-fold or greater increase in the expression of SGA-56M and/or SGA-56Mv transcript and/or protein as compared to or relative to that of a control tissue, which expresses a baseline level of SGA-56M and/or SGA-56Mv. As used herein, the terms control tissue, non-cancerous sample, or predetermined standard for a noncancerous sample may be used interchangeably to refer to a tissue that expresses a baseline level of SGA-56M and/or SGA-56Mv.

**[0064]** IMMUNOGICALLY SPECIFIC FOR: Antibodies which are immunologically specific for SGA-56M and/or SGA-56Mv are capable of recognizing SGA-56M and/or SGA-56Mv largely to the exclusion of other molecules.

[0065] SPECIFIC BINDING PAIR: A member of a specific binding pair ("sbp member") refers to one of two different molecules, having an area on the surface or in a cavity which specifically binds to and is thereby defined as complementary with a particular spatial and polar organization of another molecule. The members of the specific binding pair are referred to as ligand and receptor (antiligand). These may be members of an immunological pair such as antigen-antibody, or may be operator-repressor, nuclease-nucleotide, biotin-avidin, hormone-hormone receptor, IgG-protein A, DNA--DNA, DNA-RNA, and the like.

[0066] CONSISTING ESSENTIALLY OF: The phrase "consisting essentially of" when referring to a particular nucleotide or amino acid means a sequence having the properties of a given SEQ ID NO:. For example, when used in reference to an amino acid sequence, the phrase includes the sequence per se and molecular modifications that would not affect the basic and novel characteristics of the sequence.

[0067] PRIMER PAIR: As used herein, the terms "primer pair" or "oligonucleotide primer pair" when used in the context of a polymerase chain reaction (PCR), for example, refer to a first and a second primer of sufficient complementarity to a template nucleic acid sequence to hybridize to the template nucleic acid sequence at two physically separated sites and on separate strands such that extension from a first primer produces a single stranded nucleic acid which is at least partially complementary to a single stranded nucleic acid extended from a second primer.

[0068] PROTECTIVE IMMUNITY: As used herein, the terms "protective immunity" or "protective immune response" are intended to mean that the vaccinated subject mounts an active immune response to the antigen administered (i.e., an SGA-56M or SGA-56Mv molecule), such that upon subsequent exposure to the antigen or a cell expressing the antigen (e.g., a cancer cell), the subject is able to mount an immune response specific for the antigen or cell expressing the antigen. Such an immune response reduces the number of antigen positive cells in a subject. Thus, a protective immune response will decrease the incidence of cancer in a vaccinated subject.

[0069] FUNCTIONAL FRAGMENT: As used herein, a functional fragment of a nucleic or amino acid molecule of the invention is a fragment which retains some functional property of the larger nucleic or amino acid molecule. Examples of such functional properties include: coding for a functional polypeptide (for a nucleic acid fragment), binding to proteins, or the ability to mediate changes in cellular behavior associated with SGA-56M and/or SGA-56Mv expression, such as, for example, changes in cell morphology, cell division, differentiation, adhesion, motility, phosphorylation, or dephosphorylation of cellular proteins. One of ordinary skill in the art can readily determine using the assays described herein and those well known in the art to determine whether a fragment is a functional fragment of a nucleic or amino acid molecule using no more than routine experimentation.

[0070] The basic molecular biology techniques used to practice the methods of the invention are well known in the art, and are described for example in Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York; Ausubel et al., 1988, Current Protocols in Molecular Biology, John Wiley & Sons, New York; and Ausubel et al., 2002, Short Protocols in Molecular Biology, John Wiley & Sons, New York).

## 4. BRIEF DESCRIPTION OF THE FIGURES

[0071] FIGs. 1A and B show the nucleic acid sequence of the 2917 bp SGA-56M transcript comprising a coding sequence (CDS) spanning 225-2630bp and the amino acid sequence encoded therefrom.

[0072] FIGs. 2A and B shows the nucleic acid sequence of the 2779 bp SGA-56Mv transcript comprising a CDS spanning 225-2492bp and the amino acid sequence encoded therefrom. SGA-56Mv contains an in-frame deletion spanning 925-1063 bp of SGA-56M, as illustrated in FIG. 1.

[0073] FIG. 3 shows a photograph of an agarose gel of RT-PCR products visualized by ethidium bromide staining and generated by semi-quantitative RT-PCR of normal vs. transformed breast cells. A region common to both SGA-56M and SGA-56Mv cDNA was amplified in this assay. Samples are loaded as follows: (1) MCF-7, (2) T47D, (3) normal human mammary epithelial cells, (4) SKBR-3, (5) Hs578T, (6) MDA-MB231, (7) MDA-MB435s, (8) MDA-MB453, (9) H3396, and (10) BT549. The control gene EF-1 was included for comparison.

[0074] FIG. 4 shows a photograph of an agarose gel of RT-PCR products visualized by ethidium bromide staining and generated by semi-quantitative RT-PCR of various tumor cell-lines. A region common to both SGA-56M and SGA-56Mv cDNA was amplified in this assay. Samples are loaded as follows: (1) HCT-15, (2) HCT-116, (3) HT-29, (4) RCA, (5) NCI-H23, (6) NCI-H460, (7) NCI-H226, (8) MiaPaCa-2, (9) Bx-PC3, (10) CAPAN-2, (11) WM-115, (12) SK-MEL5, (13) SK-MEL28, (14) Colo-853, (15) Colo-857, and (16) GRM. The control gene EF-1 was included for comparison.

[0075] FIGs. 5A and B show hybridization patterns of SGA-56M/SGA-56Mv and EF-1 normal tissue expression levels on a Multiple Tissue Expression (MTE™) Array. A region common to both SGA-56M and SGA-56Mv cDNA was amplified and used as a probe for this experiment (B). The control gene EF-1 was included for comparison (A).

[0076] FIG. 6 shows a hybridization pattern of SGA-56M and SGA-56Mv on the Cancer Profiling Array (CPA™). A cancer-selective expression pattern is revealed. A region common to both SGA-56M and SGA-56Mv cDNA was amplified and used as a probe for this experiment.

[0077] FIG. 7 shows an amino acid sequence of the SGA-56M protein encoded by SEQ ID NO: 1, which comprises the CDS (SEQ ID NO: 2). A putative transmembrane region (TM) is indicated in bold and underlined spanning amino acids 340-361.

[0078] FIG. 8 shows an amino acid sequence of the SGA-56Mv protein encoded by SEQ ID NO: 3, which comprises the CDS (SEQ ID NO: 4). A putative transmembrane region (TM) is indicated in bold and underlined spanning amino acids 294-315.

[0079] FIG. 9 shows a comparison of SGA-56M and SGA-56Mv proteins with putative transmembrane regions (TMs) illustrated. SGA-56Mv includes an in-frame deletion of amino acids 234-279 of SGA-56M.

## 5. DETAILED DESCRIPTION OF THE INVENTION

[0080] The present invention relates to the discovery that the SGA-56M and SGA-56Mv genes are

over-expressed in cancer cells and tissues such as breast cancer cells. The invention relates to methods of using the SGA-56M gene and variants thereof, including SGA-56Mv, and/or the SGA-56M or SGA-56Mv gene products to diagnose, treat and prevent cancer, *e.g.*, breast cancer. The invention further relates to methods of using the SGA-56M or SGA-56Mv genes or SGA-56M or SGA-56Mv gene products to evaluate the prognosis of a patient diagnosed with cancer. The invention also relates to the discovery that the SGA-56M or SGA-56Mv genes are over-expressed in metastatic cancer cells. Thus, the invention contemplates the use of the SGA-56M gene and variants thereof, including SGA-56Mv, and/or gene products to evaluate a cancer patient's risk of metastasis of a cancer, *e.g.*, breast cancer.

**[0081]** In the development of breast neoplasia and other cancers, there is a subset of genes that are specifically expressed at various stages, and a certain number of these will be critical for the progression of malignancy, especially those associated with the metastatic spread of the disease. As described by way of example, *infra*, genes whose expression is associated with breast carcinomas at various stages of neoplastic development, were identified using Suppression Subtractive Hybridization (SSH) and high-throughput cDNA microarray analysis (Chu *et al.*, 1997, *Proc. Natl. Acad. Sci. U.S.A.* 94(19): 10057; Kuang *et al.*, 1998, *Nuc. Acids Res.* 26(4): 1116). As described herein, SSH generated cDNA libraries derived from the breast cancer cell line MCF-7 were screened using microarrays for genes which were expressed at elevated levels in the cancerous MCF-7 cells as compared to normal human mammary epithelial cells (HMECs). A total of 1536 clones were screened. The novel SGA-56M gene identified herein, and several previously identified breast cancer associated genes were identified using this approach. Details concerning the isolation and characterization of the SGA-56M cDNA and variants thereof, including SGA-56Mv, and their expression patterns in cancer cell lines and tissues is described in detail in the examples provided *infra*.

**[0082]** The present invention encompasses methods for the diagnosis, prognosis and staging of breast cancer and other cancers, *e.g.*, by the monitoring of the effect of a therapeutic treatment. Further provided are methods for the use of the SGA-56M or SGA-56Mv genes and/or SGA-56M or SGA-56Mv gene products in the identification of compounds that modulate the expression of the SGA-56M or SGA-56Mv gene or the activity of the SGA-56M or SGA-56Mv gene product. Expression of the SGA-56M gene and variants thereof including SGA-56Mv, is upregulated in various types of cancer cells including breast cancer cell lines and tissues. As such, the SGA-56M or SGA-56Mv gene products can be involved in the mechanisms underlying the onset and development of breast cancer and other types of cancer as well as the regional infiltration and metastatic spread of cancer. Thus, the present invention also provides methods for the prevention and/or treatment of breast cancer and other types of cancer, and for the control of metastatic spread of breast cancer and other types of cancer. Such methods are based on modulation of the expression and/or activity of the SGA-56M and/or SGA-56Mv gene or gene product.

**[0083]** The invention further provides for screening assays and methods of identifying agonists and antagonists of the SGA-56M or SGA-56Mv gene or gene product. The invention also provides methods of vaccinating an individual against cancer, including breast cancer, by administering an amount of the SGA-

56M or SGA-56Mv gene, gene product, or fragment thereof, in an amount that effectively elicits an immune response in a subject who has cancer or is at risk of developing cancer, including breast cancer.

## 5.1. THE SGA-56M AND SGA-56Mv GENES

**[0084]** Nucleotide sequences that encode the SGA-56M and SGA-56Mv gene open reading frames are described herein. The SGA-56M DNA (2917bp) SEQ ID NO: 1 was cloned by PCR using gene-specific primers. The SGA-56Mv DNA (2779bp) SEQ ID NO: 3 was cloned by PCR using gene-specific primers. The SGA-56M DNA sequence comprises an open reading frame SEQ ID NO: 2 spanning 225-2630bp within SEQ ID NO: 1 that encodes a protein of 802 amino acids (SEQ ID NO: 5). The SGA-56Mv DNA sequence comprises an open reading frame SEQ ID NO: 4 spanning 225-2492bp within SEQ ID NO: 3 that encodes a protein of 756 amino acids (SEQ ID NO: 6). SGA-56Mv is a variant form of SGA-56M that contains an in-frame deletion of 234-279 amino acids from within SEQ ID NO: 5. As described in detail in section 6, SGA-56M and SGA-56Mv share GenBank sequence homology with existing entries at both the nucleic acid or amino acid level using the NBLAST algorithm (www.ncbi.nlm.nih.gov).

**[0085]** The SGA-56M or SGA-56Mv nucleic acids and derivatives used in the present invention include but are not limited to RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4; a nucleic acid comprising a sequence hybridizable to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or the complement of any of the foregoing nucleic acids; a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or at least 90% homologous the complement of any of the foregoing nucleic acids (*e.g.*, as determined using the NBLAST algorithm under default parameters). As used herein an "RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4 means an RNA comprising a sequence that is the same or the (inverse) complement of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, except that thymidines (T's) can be replaced with uridines (U's). A nucleic acid derived from such RNA includes but is not limited to cDNA of said RNA, and cRNA (*e.g.*, RNA that is derived from said cDNA; *see, e.g.*, U.S. Patent Nos. 5,545,522; 5,891,636; 5,716,785). In the present invention, hybridizability can be determined under low, moderate, or high stringency conditions and preferably is under conditions of high stringency.

**[0086]** The SGA-56M or SGA-56Mv protein and derivatives used in the present invention include, but are not limited to proteins (and other molecules) comprising SEQ ID NO: 5 or SEQ ID NO: 6, proteins comprising a sequence encoded by a nucleic acid hybridizable to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4 or their complements, and proteins encoded by a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or their complement, *e.g.*, as determined using the NBLAST algorithm.

**[0087]** The SGA-56M or SGA-56Mv nucleic acids used in the present invention include but are not limited to (a) a DNA comprising the DNA sequence shown in FIG. 1 (SEQ ID NO: 1), FIG. 2 (SEQ ID NO: 3), or a complement thereof; (b) any DNA sequence that hybridizes to the DNA sequences or a complement thereof that encode the amino acid sequences shown in FIG. 7 and FIG. 8, under low, moderate or highly

stringent conditions, as disclosed *infra* in Section 5.1.1; as well as proteins encoded by such nucleic acids. In a specific embodiment, nucleic acids used in the invention encode a gene product that has at least one conservative or silent substitution. The encoded proteins are also provided for use. Additional molecules that can be used in the invention include, but are not limited to, protein derivatives that can be made by altering their sequences by substitutions, additions or deletions, and their encoding nucleic acids. Due to the degeneracy of nucleotide coding sequences, other DNA sequences that encode substantially the same amino acid sequence as a component gene or cDNA can be used in the practice of the present invention. These include but are not limited to nucleotide sequences comprising all or portions of the component protein gene that are altered by the substitution of different codons that encode a functionally equivalent amino acid residue within the sequence, thus producing a silent change. Likewise, the derivatives of the invention include, but are not limited to, those containing, as a primary amino acid sequence, part or all of the amino acid sequence of a component protein, including altered sequences in which functionally equivalent amino acid residues are substituted for residues within the sequence resulting in a conservative change. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity (a "conservative amino acid substitution") that acts as a functional equivalent, resulting in a conservative alteration. Substitutes for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

**[0088]** The invention includes the use of the SGA-56M or SGA-56Mv specific oligonucleotide sequences which preferably hybridize under highly stringent or moderately stringent conditions as described *infra* in Section 5.1.1 to at least about 6, preferably about 12, more preferably about 18, consecutive nucleotides of the SGA-56M or SGA-56Mv gene sequences described above as being useful for the detection of an SGA-56M or SGA-56Mv gene product for the diagnosis and prognosis of cancer. Such gene products include, *e.g.*, an RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4; a nucleic acid comprising a sequence hybridizable to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, or its complement under conditions of high stringency; a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4 or its complement as determined using the NBLAST algorithm.

**[0089]** The invention also includes the use of nucleic acid molecules, preferably DNA molecules, that preferably hybridize under highly stringent or moderately stringent conditions as described *infra* in Section 5.1.1 to, and are therefore the inverse complements of, the nucleic acid sequences, described, *inter alia*, in Section 3 above. These nucleic acid molecules may encode or act as antisense molecules that may be used to partially or completely inhibit SGA-56M and/or SGA-56Mv expression. With respect to SGA-56M or SGA-56Mv gene regulation, such techniques can be used to modulate, for example, the phenotype and

metastatic potential of breast cancer or other cancer cells. Moreover, such sequences also provide components of ribozyme and/or triple helix sequences that may be used to advantage to modulate (e.g., inhibit) SGA-56M or SGA-56Mv gene expression. Thus, these sequences and ribozyme and/or triple helix sequences comprising such sequences may be used for the treatment and/or prevention of cancer.

**[0090]** In one embodiment, the invention encompasses methods of using the SGA-56M or SGA-56Mv gene coding sequence or fragments and degenerate variants of DNA sequences which encode the SGA-56M or SGA-56Mv gene or gene product, including naturally occurring and non-naturally occurring variants thereof. A non-naturally occurring variant is one that is engineered by man. A naturally occurring SGA-56M or SGA-56Mv gene, gene product, or variant thereof is one that is not engineered by man. In the methods of the invention wherein an SGA-56M or SGA-56Mv gene product in a sample derived from a subject is detected or measured, naturally occurring SGA-56M or SGA-56Mv gene products are detected, including, but not limited to wild-type SGA-56M or SGA-56Mv gene products as well as mutants, allelic variants, splice variants, polymorphic variants, *etc.* In general, such mutants and variants are believed to be highly homologous to SEQ ID NO: 1, or SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4, *e.g.*, at least 90% homologous and/or hybridizable under high stringency conditions. In specific embodiments, the mutants and variants being detected or measured may comprise (or, if nucleic acids, encode) not more than 1, 2, 3, 4, or 5 point mutations (substitutions) relative to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4 and/or comprise or encode only conservative amino acid substitutions.

**[0091]** In other methods of the invention, wild-type, or naturally occurring variant, or non-naturally occurring variant SGA-56M or SGA-56Mv sequences may be used in the methods of the invention (*e.g.*, in vaccination, immunization, antisense, or ribozyme procedures).

**[0092]** An SGA-56M or SGA-56Mv gene fragment may be a complementary DNA (cDNA) molecule or a genomic DNA molecule that may comprise one or more intervening sequences or introns, as well as regulating regions located beyond the 5' and 3' ends of the coding region or within an intron.

**[0093]** The present invention provides for methods of using isolated nucleic acid molecules encoding an SGA-56M or SGA-56Mv protein, polypeptide, or fragments, derivatives, and variants thereof that include, both naturally occurring and non-naturally occurring variants or mutants. The invention also contemplates, for use in the methods of the invention, the use of 1) any nucleic acid that encodes an SGA-56M or SGA-56Mv polypeptide of the invention; 2) any nucleic acid that hybridizes to the complement of the sequences disclosed herein, preferably under highly stringent conditions as disclosed *infra* in Section 5.1.1, and encodes a functionally equivalent gene product; and/or 3) any nucleic acid sequence that hybridizes to the complement of the sequences disclosed herein, preferably under moderately stringent conditions, as disclosed *infra* in Section 5.1.1 yet which still encodes a gene product that displays a functional activity of SGA-56M or SGA-56Mv.

**[0094]** As discussed above, the invention also contemplates the use of isolated nucleic acid molecules that encode a variant protein or polypeptide. The variant protein or polypeptide can occur naturally or non-naturally. It can be engineered by introducing nucleotide substitutions, *e.g.*, point mutations, or

additions or deletions into the nucleotide sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4. In a specific embodiment, one or more, but not more than 5, 10, or 25 amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

**[0095]** In a specific embodiment, the invention provides for the use of SGA-56M or SGA-56Mv derivatives and analogs of the invention which are functionally active, *i.e.*, they are capable of displaying one or more known functional activities associated with a (wild-type) SGA-56M or SGA-56Mv-encoded protein. Such functional activities include but are not limited to antigenicity/immunogenicity (ability to bind or compete with SGA-56M or SGA-56Mv for binding to an anti-SGA-56M or anti-SGA-56Mv antibody, respectively or ability to generate antibody which binds to SGA-56M or SGA-56Mv), ability to bind or compete with SGA-56M or SGA-56Mv for binding to other proteins or fragments thereof, such as proteins capable of forming complexes with SGA-56M and/or SGA-56Mv.

**[0096]** The nucleic acid sequences of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4 or portions thereof, may be used, for example, as hybridization probes. Nucleic acid molecules encoding an SGA-56M or SGA-56Mv gene product can be isolated using standard hybridization and cloning techniques (*See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989 for methodological details pertaining to the methods of the invention.

**[0097]** In addition, gene products encoded by SGA-56M or SGA-56Mv, including SGA-56M or SGA-56Mv peptide fragments can comprise components of fusion proteins, which may be used to facilitate recovery, detection, and/or localization of another protein of interest. Antibodies immunologically specific for SGA-56M and/or SGA-56Mv may also be used effectively to recover, detect, and/or localize another protein of interest that, for example, binds to SGA-56M and/or SGA-56Mv. In addition, genes and gene products encoded for by SGA-56M or SGA-56Mv can be used as research reagents, *e.g.*, for genetic mapping.

**[0098]** Additionally, the present invention contemplates use of the nucleic acid molecules, polypeptides, and/or antagonists of gene products encoded for by the SGA-56M or SGA-56Mv gene to screen, diagnose, prevent and/or treat disorders characterized by aberrant expression or activity of the SGA-56M or SGA-56Mv polypeptides, which include, cancers, such as, but not limited to cancer of the breast, ovary, skin and lymphoid system.

**[0099]** The present invention encompasses the use of SGA-56M or SGA-56Mv nucleic acid molecules comprising cDNA, genomic DNA, introns, exons, promoter regions, 5' and 3' regulatory regions of the gene, RNA, hnRNA, mRNA, regulatory regions within RNAs, and degenerate variants thereof in the methods of the invention. Promoter sequences for SGA-56M or SGA-56Mv can be determined by promoter-reporter gene assays and *in vitro* binding assays.

**[0100]** In one embodiment, the invention comprises the use of a variant SGA-56M or SGA-56Mv nucleic acid sequence that hybridizes to a naturally-occurring occurring variant SGA-56M or SGA-56Mv nucleic acid molecule under stringent conditions as described *infra* in Section 5.1.1. In another embodiment, the invention contemplates the use of an SGA-56M or SGA-56Mv variant nucleic acid sequence that hybridizes to a naturally-occurring occurring variant SGA-56M or SGA-56Mv nucleic acid molecule under moderately stringent conditions as described *infra* in Section 5.1.1.

**[0101]** A nucleic acid molecule is intended to include DNA molecules (*e.g.*, cDNA, genomic DNA), RNA molecules (*e.g.*, hnRNA, pre-mRNA, mRNA), and DNA or RNA analogs generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded.

**[0102]** The SGA-56M or SGA-56Mv gene sequences used in the methods of the invention are of human origin, however, homologs of SGA-56M or SGA-56Mv isolated from other mammals may also be used in the methods of the invention. Thus, the invention also includes the use of SGA-56M or SGA-56Mv homologs isolated from non-human animals such as: non-human primates; rats; mice; farm animals including, but not limited to: cattle; horses; goats; sheep; pigs; *etc.*; household pets including, but not limited to: cats; dogs; *etc.* in the methods of the invention.

**[0103]** Still further, such molecules may be used as components of diagnostic and/or prognostic methods whereby, for example, the presence of a particular SGA-56M or SGA-56Mv allele or alternatively spliced SGA-56M or SGA-56Mv transcript responsible for causing or predisposing one to breast cancer or other cancers may be detected.

**[0104]** The invention also includes the use of transcriptional regulators that control the level of expression of an SGA-56M or SGA-56Mv gene product. A transcriptional regulator can include, *e.g.*, a protein that binds a DNA sequence and up-regulates or down-regulates the transcription of the SGA-56M or SGA-56Mv gene. A transcriptional regulator can also include a nucleic acid sequence that can be either upstream or downstream from the SGA-56M or SGA-56Mv gene and which binds an effector molecule that enhances or suppresses SGA-56M or SGA-56Mv gene transcription.

**[0105]** Still further, the invention encompasses the use of SGA-56M or SGA-56Mv gene coding sequences or fragments thereof in screens to identify proteins, peptides or nucleic acids related to the onset and/or metastatic spread of cancer, including breast and lung cancer. Examples of engineered yeast-based systems for investigating protein-protein interactions include, but are not limited to, the yeast two-hybrid system.

**[0106]** The invention also encompasses the use of (a) DNA vectors comprising any of the foregoing SGA-56M or SGA-56Mv coding sequences and/or their complements (*e.g.*, antisense); (b) DNA expression vectors comprising any of the foregoing SGA-56M or SGA-56Mv coding sequences operatively linked or associated with a regulatory element that directs the expression of the coding sequences; and (c) genetically engineered host cells comprising any of the foregoing SGA-56M or SGA-56Mv coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences in the host cell. Cell lines and/or vectors which comprise and/or express SGA-56M or SGA-56Mv can be used to produce an SGA-56M or SGA-56Mv gene product for use in the methods of the invention. Such methods include, *e.g.*, vaccination

against breast cancer or other cancers in which expression of SGA-56M or SGA-56Mv is found to be elevated and screening assays to identify antagonists and agonists that bind and/or interact with SGA-56M and/or SGA-56Mv or modulate (i.e., suppress or enhance) expression of SGA-56M and/or SGA-56Mv.

**[0107]** As used herein, regulatory elements include, but are not limited to inducible and non-inducible promoters, enhancers, tissue specific promoters and/or enhancers, operators and other elements that drive and regulate expression and are known to those skilled in the art. Such regulatory elements include but are not limited to the cytomegalovirus (hCMV) immediate early promoter, the early or late promoters of SV40 adenovirus, the lac system, the trp system, the TAC system, the TRC system, the major operator and promoter regions of phage A, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase, the promoters of acid phosphatase, and the promoters of the yeast α-mating factors.

**[0108]** The invention includes the use of fragments or derivatives of any of the nucleic acids disclosed herein in any of the methods of the invention. In various embodiments, a fragment or derivative comprises 10, 20, 50, 100, 200, or more nucleotides of SEQ ID NO: 1 or SEQ ID NO: 3.

**[0109]** In addition to the use of the SGA-56M or SGA-56Mv gene sequences described above, homologs of such sequences, exhibiting extensive homology to the SGA-56M or SGA-56Mv gene product present in other species can be identified and readily isolated, and used in the methods of the invention without undue experimentation, by molecular biological techniques well known in the art. Further, there can exist homolog genes at other genetic loci within the genome that encode proteins that have extensive homology to SGA-56M or SGA-56Mv. Such homologous genes, can encode multiple proteins, one or both of which are homologous to SGA-56M and/or SGA-56Mv. Alternatively, such homologous genes can encode a single protein with homology to SGA-56M and/or SGA-56Mv. Once identified, such genes can be used in the methods of the present invention. Still further, there can exist alternatively spliced variants of the SGA-56M or SGA-56Mv gene. The invention thus includes the use of any of homolog/ortholog/variant of SGA-56M and/or SGA-56Mv in the methods of the invention.

**[0110]** As an example, in order to clone a mammalian SGA-56M or SGA-56Mv gene homolog or variants using isolated human SGA-56M or SGA-56Mv gene sequences as disclosed herein, such human SGA-56M or SGA-56Mv gene sequences are labeled and used to screen a cDNA library constructed from mRNA obtained from appropriate cells or tissues (*e.g.*, breast epithelial cells) derived from the organism of interest. With respect to the cloning of such a mammalian SGA-56M or SGA-56Mv homolog, a mammalian breast cancer cell cDNA library may, for example, be used for screening. In one embodiment, such a screen would employ a probe corresponding to all or a portion of the SGA-56M or SGA-56Mv open reading frame SEQ ID NO: 2 or SEQ ID NO: 4. In yet another embodiment, such a screen would employ one or more probes corresponding to all or a portion of the coding sequence for SGA- 56M (SEQ ID NO: 2) or SGA-56Mv (SEQ ID NO: 4), for example, a probe corresponding to the SGA-56M (SEQ ID NO: 1) or SGA-56Mv (SEQ ID NO: 3).

[0111] The hybridization and wash conditions used should be of a low stringency, as described *infra* in Section 5.1.1 when the cDNA library is derived from a species other than the species from which the labeled sequence (e.g., probe) was derived.

[0112] Alternatively, the labeled fragment (or probe) may be used to screen a genomic library derived from the organism of interest, again, using appropriately stringent conditions well known to those of skill in the art.

[0113] Further, an SGA-56M or SGA-56Mv gene homolog/ortholog/variant may be isolated from nucleic acid of an organism of interest by performing PCR using two degenerate oligonucleotide primer pools based on amino acid sequences of SGA-56M and/or SGA-56Mv encoded gene products. The template for the reaction may, for example, be cDNA obtained by reverse transcription of mRNA prepared from, for example, mammalian cell lines or tissue known or suspected to express an allele, homolog, ortholog, or variant of SGA-56M and/or SGA-56Mv.

[0114] The PCR product may be subcloned and sequenced to ensure that the amplified sequences represent the sequences of an SGA-56M or SGA-56Mv-related nucleic acid sequence. The PCR fragment may then be used to isolate a larger fragment of an SGA-56M or SGA-56Mv-related nucleic acid sequence (e.g., a full length cDNA clone) by a variety of methods. For example, the amplified fragment may be labeled and used to screen a cDNA library, such as a bacteriophage cDNA library. Alternatively, the labeled fragment may be used to isolate genomic clones via the screening of a genomic library.

[0115] PCR technology may be utilized to isolate additional nucleic acid sequences. For example, RNA may be isolated, following standard procedures, from an appropriate cellular or tissue source (*e.g.*, one known, or suspected, to express the SGA-56M or SGA-56Mv gene, such as, for example, breast cancer cell-lines). A reverse transcription reaction may be performed on the RNA using an oligonucleotide primer specific or selective for the most 5' end of the amplified fragment for the priming of first strand synthesis. The resulting RNA/DNA hybrid may then be "tailed" with guanines using a standard terminal transferase reaction, the hybrid may be digested with RNAase H, and second strand synthesis may then be primed with a poly-C primer. Thus, nucleic acid sequences upstream of the amplified fragment may easily be isolated. For a review of PCR technology and cloning strategies which may be used, see, *e.g.*, PCR Primer, 1995, Dieffenbach *et al.*, ed., Cold Spring Harbor Laboratory Press; Sambrook *et al.*, 1989, *supra.*

[0116] SGA-56M or SGA-56Mv gene coding sequences may additionally be used to isolate mutant SGA-56M or SGA-56Mv gene alleles. Such mutant alleles may be isolated from individuals either known or susceptible to or predisposed to have a genotype that contributes to the development of cancer, *e.g.*, breast cancer, including metastasis. Such mutant alleles may also be isolated from individuals either known or susceptible to or predisposed to have a genotype that contributes to resistance to the development of cancer, *e.g.*, breast cancer, including metastasis. Mutant alleles and mutant allele products may then be utilized in the screening, therapeutic and diagnostic methods and systems described herein. Additionally, such SGA-56M or SGA-56Mv gene sequences can be used to detect SGA-56M or SGA-56Mv gene regulatory (*e.g.*, promoter) defects that can affect the development and outcome of cancer. Mutants can be isolated by any technique known in the art, *e.g.*, PCR, screening genomic libraries, screening expression libraries.

[0117] As described below, the invention also relates to the use of an SGA-56M or SGA-56Mv gene coding sequence or gene product in the methods of the invention. An SGA-56M or SGA-56Mv gene coding sequence or gene product includes, but is not limited to an RNA corresponding to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, a protein comprising SEQ ID NO: 5 or SEQ ID NO: 6, or a nucleic acid comprising a sequence hybridizable to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4 under conditions of high stringency, or a protein comprising a sequence encoded by said hybridizable sequence or a nucleic acid at least 90% homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4 as determined using the NBLAST algorithm or a protein encoded thereby.

### 5.1.1 HYBRIDIZATION CONDITIONS

[0118] A nucleic acid which is hybridizable to an SGA-56M or SGA-56Mv nucleic acid (*e.g.*, having a sequence as set forth in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4 or to its reverse complement), or to a nucleic acid encoding an SGA-56M or SGA-56Mv derivative, or to its reverse complement under conditions of low stringency can be used in the methods of the invention to detect the presence of an SGA-56M or SGA-56Mv gene and/or presence or expression level of an SGA-56M or SGA-56Mv gene product. By way of example and not limitation, procedures using such conditions of low stringency are as follows (*see also* Shilo and Weinberg, 1981, Proc. Natl. Acad. Sci. U.S.A. 78, 6789-6792). Filters containing DNA are pretreated for 6 h at 40°C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 µg/ml denatured salmon sperm DNA. Hybridizations are carried out in the same solution with the following modifications: 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 µg/ml salmon sperm DNA, 10% (wt/vol) dextran sulfate, and 5-20 X $10^6$ cpm $^{32}$P-labeled probe is used. Filters are incubated in hybridization mixture for 18-20 h at 40°C, and then washed for 1.5 h at 55°C in a solution containing 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS. The wash solution is replaced with fresh solution and incubated an additional 1.5 h at 60°C. Filters are blotted dry and exposed for autoradiography. If necessary, filters are washed for a third time at 65-68°C and re-exposed to film. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations).

[0119] A nucleic acid which is hybridizable to an SGA-56M or SGA-56Mv nucleic acid (*e.g.*, having a sequence as set forth in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4 or to its reverse complement), or to a nucleic acid encoding an SGA-56M or SGA-56Mv derivative, or to its reverse complement under conditions of high stringency is also provided for use in the methods of the invention. By way of example and not limitation, procedures using such conditions of high stringency are as follows. Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X $10^6$ cpm of $^{32}$P-labeled probe. Washing of filters is done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA. This is followed

by a wash in 0.1X SSC at 50°C for 45 min before autoradiography. Other conditions of high stringency that may be used are well known in the art.

**[0120]** A nucleic acid which is hybridizable to an SGA-56M or SGA-56Mv nucleic acid (*e.g.*, having a sequence as set forth in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4 or to its reverse complement), or to a nucleic acid encoding an SGA-56M or SGA-56Mv derivative, or to its reverse complement under conditions of moderate stringency is also provided for use in the methods of the invention. For example, but not limited to, procedures using such conditions of moderate stringency may be performed according to the following method. Filters containing DNA are pretreated for 6 hours at 55°C in a solution containing 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 μg/ml denatured salmon sperm DNA. Hybridizations are carried out in the same solution with 5-20 x $10^6$ cpm $^{32}$P-labeled probe. Filters are incubated in hybridization mixture for 18-20 hours at 55°C, and then washed twice for 30 minutes at 60°C in a solution containing 1X SSC and 0.1% SDS. Filters are blotted dry and exposed for autoradiography. Washing of filters is done at 37°C for 1 hour in a solution containing 2X SSC, 0.1% SDS. Other conditions of moderate stringency that may be used are well known in the art. (see, *e.g.*, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; *see also*, Ausubel et al., eds., in the Current Protocols in Molecular Biology series of laboratory technique manuals, 1987-1997 Current Protocols,© 1994-1997 John Wiley and Sons, Inc.).

## 5.2. PROTEIN PRODUCTS OF THE SGA-56M AND SGA-56Mv GENES

**[0121]** In another embodiment, the present invention provides for the use of SGA-56M or SGA-56Mv gene products, including SGA-56M or SGA-56Mv, and/or peptide fragments thereof which can be used for the generation of antibodies, in diagnostic assays, or for the identification of other cellular gene products involved in the development of cancer, such as, for example, breast cancer.

**[0122]** The amino acid sequences depicted in FIG. 7 and FIG. 8 represent examples of SGA-56M or SGA-56Mv gene products, *i.e.*, SGA-56M (SEQ ID NO: 5) or SGA-56Mv (SEQ ID NO: 6). The SGA-56M or SGA-56Mv gene products, sometimes referred to herein as an "SGA-56M or SGA-56Mv proteins" or "SGA-56M or SGA-56Mv polypeptides," may additionally include those gene products encoded by the SGA-56M or SGA-56Mv gene sequences described in Section 5.1, above.

**[0123]** In addition, SGA-56M or SGA-56Mv derivatives may include proteins that have conservative amino acid substitution(s) and/or display a functional activity of an SGA-56M or SGA-56Mv gene product, including but not limited to SGA-56M or SGA-56Mv. Such a derivative may contain deletions, additions or substitutions of amino acid residues within the amino acid sequence encoded by the SGA-56M or SGA-56Mv gene sequences described, above, in Section 5.1, but which result in a conservative change, thus producing a functionally equivalent SGA-56M or SGA-56Mv gene product.

**[0124]** In a specific embodiment, the invention provides a functionally equivalent protein that exhibits a substantially similar *in vivo* activity as an endogenous SGA-56M or SGA-56Mv gene product encoded by an SGA-56M or SGA-56Mv gene sequence described in Section 5.1, above. An *in vivo* activity of the SGA-56M

or SGA-56Mv gene product can be exhibited by, for example, preneoplastic and/or neoplastic transformation of a cell upon overexpression of the gene product, such as for example, may occur in the onset and progression and metastasis of breast cancer.

**[0125]** An SGA-56M or SGA-56Mv gene product sequence preferably comprises an amino acid sequence that exhibits at least about 65% sequence similarity to SGA-56M or SGA-56Mv, more preferably exhibits at least 70% sequence similarity to SGA-56M or SGA-56Mv, yet more preferably exhibits at least about 75% sequence similarity to SGA-56M or SGA-56Mv. In other embodiments, the SGA-56M or SGA-56Mv gene product sequence preferably comprises an amino acid sequence that exhibits at least 85% sequence similarity to SGA-56M or SGA-56Mv, yet more preferably exhibits at least 90% sequence similarity to SGA-56M or SGA-56Mv, and most preferably exhibits at least about 95% sequence similarity to SGA-56M or SGA-56Mv.

**[0126]** The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) *Proc Natl Acad Sci.* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc Natl Acad Sci.* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402. Alternatively, PSI-Blast can be used to perform an iterated search that detects distant relationships between molecules (*Id.*). When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. *See* http://www.ncbi.nlm.nih.gov.

**[0127]** Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, (1988) CABIOS 4:11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art and include ADVANCE and ADAM as described in Torellis and Robotti (1994) *Comput. Appl. Biosci.*, 10:3-5; and FASTA described in Pearson and Lipman (1988) 85:2444-8. Within FASTA, ktup is a control option that sets the sensitivity and speed of the search. If ktup=2, similar regions in the two sequences being compared are found by looking at pairs of aligned residues; if ktup=1, single aligned amino acids are examined. ktup can be set to 2 or 1 for protein sequences, or from 1 to 6 for DNA sequences. The default if ktup is not specified is 2 for proteins and 6 for DNA. For a further description of FASTA parameters, see http://bioweb.pasteur.fr/docs/man/man/fasta.1.html#sect2.

**[0128]** The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, only exact matches are

counted. However, conservative substitutions should be considered in evaluating sequences that have a low percent identity with the SGA-56M or SGA-56Mv sequences disclosed herein.

**[0129]** In a specific embodiment, molecules or protein comprising at least 10, 20, 30, 40, 50, 75, 100, or 200 amino acids of SEQ ID NO: 5 or SEQ ID NO: 6 are used in the present invention.

### 5.2.1 FUSION PROTEINS

**[0130]** SGA-56M or SGA-56Mv gene products can also include fusion proteins comprising an SGA-56M or SGA-56Mv gene product sequence as described above operatively linked or associated to a heterologous, component, *e.g.*, peptide for use in the methods of the invention. Heterologous components can include, but are not limited to sequences that facilitate isolation and purification of fusion protein, or label components. Heterologous components can also include sequences that confer stability to the SGA-56M or SGA-56Mv gene product or target the gene product to, for example, a particular tissue or cell type. Such isolation, labeling, and targeting components are well known to those of skill in the art.

**[0131]** The present invention encompasses the use of fusion proteins comprising the protein or fragment thereof encoded for by the SGA-56M or SGA-56Mv gene open reading frames SEQ ID NO: 2 or SEQ ID NO: 4 and a heterologous polypeptide (*i.e.*, an unrelated polypeptide or fragment thereof, preferably at least 10 to 100 amino acids of the polypeptide). The fusion can be direct, but may occur through linker sequences. The heterologous polypeptide may be fused to the N-terminus or C-terminus of an SGA-56M or SGA-56Mv gene product.

**[0132]** A fusion protein can comprise an SGA-56M or SGA-56Mv gene product fused to a signal sequence at its N-terminus. Various signal sequences are commercially available. Eukaryotic heterologous signal sequences include, but are not limited to, the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). Prokaryotic heterologous signal sequences useful in the methods of the invention include, but are not limited to, the phoA secretory signal (Sambrook *et al.*, eds., *Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

**[0133]** The SGA-56M or SGA-56Mv protein (SEQ ID NO: 5 or SEQ ID NO: 6) or fragment thereof encoded by the SGA-56M or SGA-56Mv open reading frames SEQ ID NO: 2 or SEQ ID NO: 4, respectively, can be fused to tag sequences, *e.g.*, a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., Chatsworth, CA, 91311), among others, many of which are commercially available for use in the methods of the invention. As described in Gentz *et al.*, 1989, *Proc. Natl. Acad. Sci. USA*, 86:821-824, for instance, hexa-histidine provides for convenient purification of the fusion protein. Other examples of peptide tags are the hemagglutinin "HA" tag, which corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson *et al.*, 1984, *Cell*, 37:767) and the "flag" tag (Knappik *et al.*, 1994, *Biotechniques*, 17(4):754-761). These tags are especially useful for purification of recombinantly produced polypeptides of the invention.

**[0134]** A fusion protein may readily be purified utilizing an antibody specific/selective for the fusion protein being expressed. For example, a system described by Janknecht *et al.* allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (Janknecht *et al.*, 1991, *Proc. Natl. Acad. Sci. USA* 88:8972). In this system, the gene of interest is subcloned into a vaccinia recombination plasmid such that the open reading frame of the gene is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto $Ni^{2+}$·nitriloacetic acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

**[0135]** An affinity label may be fused, for example, at either the amino or carboxyl terminal of the protein or fragment thereof encoded by an SGA-56M or SGA-56Mv open reading frame to generate a fusion protein for use in the methods of the invention. The precise site at which a fusion is made in the carboxyl terminal, for example, is not critical. The optimal site can be determined by routine experimentation.

**[0136]** A variety of affinity labels known in the art may be used, such as, but not limited to, the immunoglobulin constant regions, (Petty, 1996, Metal-chelate affinity chromatography, in Current Protocols in Molecular Biology, Vol. 2, Ed. Ausubel *et al.*, Greene Publish. Assoc. & Wiley Interscience), glutathione S-transferase (GST; Smith, 1993, *Methods Mol. Cell Bio.* 4:220-229), the *E. coli* maltose binding protein (Guan *et al.*, 1987, *Gene* 67:21-30), and various cellulose binding domains (U.S. Patent Nos. 5,496,934; 5,202,247; 5,137,819; Tomme *et al.*, 1994, *Protein Eng.* 7:117-123), etc. Other affinity labels may impart fluorescent properties to an SGA-56M or SGA-56Mv gene product, *e.g.*, green fluorescent protein and the like. Other affinity labels are recognized by specific binding partners and thus facilitate isolation by affinity binding to the binding partner that can be immobilized onto a solid support. Some affinity labels may afford the SGA-56M or SGA-56Mv gene product novel structural properties, such as the ability to form multimers. These affinity labels are usually derived from proteins that normally exist as homopolymers. Affinity labels such as the extracellular domains of CD8 (Shiue *et al.*, 1988, *J. Exp. Med.* 168:1993-2005), or CD28 (Lee *et al.*, 1990, *J. Immunol.* 145:344-352), or fragments of the immunoglobulin molecule containing sites for interchain disulfide bonds, could lead to the formation of multimers.

**[0137]** As will be appreciated by those skilled in the art, many methods can be used to obtain the coding region of the above-mentioned affinity labels, including but not limited to, DNA cloning, DNA amplification, and synthetic methods. Some of the affinity labels and reagents for their detection and isolation are available commercially.

**[0138]** A preferred affinity label is a non-variable portion of the immunoglobulin molecule. Typically, such portions comprise at least a functionally operative CH2 and CH3 domain of the constant region of an immunoglobulin heavy chain. Fusions are also made using the carboxyl terminus of the Fc portion of a constant domain, or a region immediately amino-terminal to the CH1 of the heavy or light chain. Suitable immunoglobulin-based affinity label may be obtained from IgG-1, -2, -3, or -4 subtypes, IgA, IgE, IgD, or IgM, but preferably IgG1. Preferably, a human immunoglobulin is used when the SGA-56M or SGA-56Mv gene product is intended for *in vivo* use for humans. Many DNA encoding immunoglobulin light or heavy

chain constant regions are known or readily available from cDNA libraries. See, for example, Adams *et al.*, *Biochemistry*, 1980, 19:2711-2719; Gough *et al.*, 1980, *Biochemistry*, 19:2702-2710; Dolby *et al.*, 1980, *Proc. Natl. Acad. Sci. U.S.A.*, 77:6027-6031; Rice *et al.*, 1982, *Proc. Natl. Acad. Sci. U.S.A.*, 79:7862-7865; Falkner *et al.*, 1982, *Nature*, 298:286-288; and Morrison *et al.*, 1984, *Ann. Rev. Immunol*, 2:239-256. Because many immunological reagents and labeling systems are available for the detection of immunoglobulins, the SGA-56M or SGA-56Mv gene product-Ig fusion protein can readily be detected and quantified by a variety of immunological techniques known in the art, such as the use of enzyme-linked immunosorbent assay (ELISA), immunoprecipitation, fluorescence activated cell sorting (FACS), etc. Similarly, if the affinity label is an epitope with readily available antibodies, such reagents can be used with the techniques mentioned above to detect, quantitate, and isolate the SGA-56M or SGA-56Mv gene product containing the affinity label. In many instances, there is no need to develop specific or selective antibodies to the SGA-56M or SGA-56Mv gene product.

**[0139]** A fusion protein can comprise an SGA-56M or SGA-56Mv gene product fused to the Fc domain of an immunoglobulin molecule or a fragment thereof for use in the methods of the invention. A fusion protein can also comprise an SGA-56M or SGA-56Mv gene product fused to the CH2 and/or CH3 region of the Fc domain of an immunoglobulin molecule. Furthermore, a fusion protein can comprise an SGA-56M or SGA-56Mv gene product fused to the CH2, CH3, and hinge regions of the Fc domain of an immunoglobulin molecule (see Bowen *et al.*, 1996, *J. Immunol.* 156:442-49). This hinge region contains three cysteine residues that are normally involved in disulfide bonding with other cysteines in the Ig molecule. Since none of the cysteines are required for the peptide to function as a tag, one or more of these cysteine residues may optionally be substituted by another amino acid residue, such as for example, serine.

**[0140]** Various leader sequences known in the art can be used for the efficient secretion of the SGA-56M or SGA-56Mv gene product from bacterial and mammalian cells (von Heijne, 1985, J. *Mol. Biol.* 184:99-105). Leader peptides are selected based on the intended host cell, and may include bacterial, yeast, viral, animal, and mammalian sequences. For example, the herpes virus glycoprotein D leader peptide is suitable for use in a variety of mammalian cells. A preferred leader peptide for use in mammalian cells can be obtained from the V-J2-C region of the mouse immunoglobulin kappa chain (Bernard *et al.*, 1981, *Proc. Natl. Acad. Sci.* 78:5812-5816). Preferred leader sequences for targeting SGA-56M or SGA-56Mv gene product expression in bacterial cells include, but are not limited to, the leader sequences of the *E. coli* proteins OmpA (Hobom *et al.*, 1995, *Dev. Biol. Stand.* 84:255-262), Pho A (Oka *et al.*, 1985, *Proc. Natl. Acad. Sci* 82:7212-16), OmpT (Johnson *et al.*, 1996, *Protein Expression* 7:104-113), LamB and OmpF (Hoffman & Wright, 1985, *Proc. Natl. Acad. Sci. USA* 82:5107-5111), β-lactamase (Kadonaga *et al.*, 1984, *J. Biol. Chem.* 259:2149-54), enterotoxins (Morioka-Fujimoto *et al.*, 1991, *J. Biol. Chem.* 266:1728-32), *Staphylococcus aureus* protein A (Abrahmsen *et al.*, 1986, *Nucleic Acids Res.* 14:7487-7500), and the *B. subtilis* endoglucanase (Lo *et al.*, *Appl. Environ. Microbiol.* 54:2287-2292), as well as artificial and synthetic signal sequences (MacIntyre *et al.*, 1990, *Mol. Gen. Genet.* 221:466-74; Kaiser *et al.*, 1987, *Science*, 235:312-317).

**[0141]** A fusion protein can comprise an SGA-56M or SGA-56Mv gene product and a cell permeable peptide, which facilitates the transport of a protein or polypeptide across the plasma membrane for use in the methods of the invention. Examples of cell permeable peptides include, but are not limited to, peptides derived from hepatitis B virus surface antigens (*e.g.*, the PreS2- domain of hepatitis B virus surface antigens), herpes simplex virus VP22, antennapaedia, 6H, 6K, and 6R. See, *e.g.*, Oess *et al.*, 2000, *Gene Ther.* 7:750-758, DeRossi *et al.*, 1998, *Trends Cell Biol* 8(2):84-7, and Hawiger, 1997, *J. Curr Opin Immunol* 9(2):189-94.

**[0142]** Fusion proteins can be produced by standard recombinant DNA techniques or by protein synthetic techniques, *e.g.*, by use of a peptide synthesizer. For example, a nucleic acid molecule encoding a fusion protein can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g., Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., John Wiley & Sons, 1992).

**[0143]** The nucleotide sequence coding for a fusion protein can be inserted into an appropriate expression vector, *i.e.*, a vector that contains the necessary elements for the transcription and translation of the inserted protein-coding sequence. The expression of a fusion protein may be regulated by a constitutive, inducible or tissue-specific, or selective promoter. It will be understood by the skilled artisan that fusion proteins, which can facilitate solubility and/or expression, or can increase the *in vivo* half-life of the protein or fragment thereof encoded by SGA-56M (SEQ ID NO: 2) or SGA-56Mv (SEQ ID NO: 4) and thus are useful in the methods of the invention. The SGA-56M or SGA-56Mv gene products or peptide fragments thereof, or fusion proteins can be used in any assay that detects or measures SGA-56M or SGA-56Mv gene products or in the calibration and standardization of such assay.

**[0144]** The methods of the invention encompass the use of SGA-56M or SGA-56Mv gene products or peptide fragments thereof, which may be produced by recombinant DNA technology using techniques well known in the art. Thus, methods for preparing the SGA-56M or SGA-56Mv gene polypeptides and peptides of the invention by expressing nucleic acid containing SGA-56M or SGA-56Mv gene sequences are described herein. Methods that are well known to those skilled in the art can be used to construct expression vectors containing SGA-56M or SGA-56Mv gene product coding sequences (including but not limited to SEQ ID NO: 2 or SEQ ID NO: 4 and appropriate transcriptional and translational control signals. These methods include, for example, *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. See, for example, the techniques described in Sambrook *et al.*, 1989, *supra*, and Ausubel *et al.*, 1989, *supra*. Alternatively, RNA capable of encoding SGA-56M or SGA-56Mv gene product sequences may be chemically synthesized using, for example, synthesizers (*see e.g.*, the techniques described in *Oligonucleotide Synthesis*, 1984, Gait, M.J. ed., IRL Press, Oxford).

5.2.2 EXPRESSION SYSTEMS

[0145] A variety of host-expression vector systems may be utilized to express the SGA-56M or SGA-56Mv gene coding sequences for use in the methods of the invention. Such host-expression systems represent vehicles by which the coding sequences of interest may be produced and subsequently purified, but also represent cells which may, when transformed or transfected with the appropriate nucleotide coding sequences, exhibit the SGA-56M or SGA-56Mv gene product of the invention *in situ*. These include but are not limited to microorganisms such as bacteria (*e.g.*, *E. coli*, *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing SGA-56M or SGA-56Mv gene product coding sequences; yeast (*e.g.*, *Saccharomyces*, *Pichia*) transformed with recombinant yeast expression vectors containing the SGA-56M or SGA-56Mv gene product coding sequences; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing the SGA-56M or SGA-56Mv gene product coding sequences; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; or tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.*, Ti plasmid) containing SGA-56M or SGA-56Mv gene product coding sequences; or mammalian cell systems (*e.g.*, COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5K promoter).

[0146] In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the SGA-56M or SGA-56Mv gene product being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of pharmaceutical compositions of SGA-56M or SGA-56Mv protein or for raising antibodies to SGA-56M or SGA-56Mv protein, vectors that direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the *E. coli* expression vector pUR278 (Ruther *et al.*, 1983, *EMBO J.* 2:1791), in which the SGA-56M or SGA-56Mv gene product coding sequence may be ligated individually into the vector in frame with the *lac Z* coding region so that a fusion protein is produced; pIN vectors (Inouye & Inouye, 1985, *Nucleic Acids Res.* 13:3101; Van Heeke & Schuster, 1989, *J. Biol. Chem.* 264:5503); and the like. pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption and binding to a matrix of glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include, *e.g.*, thrombin or factor Xa protease cleavage sites so that the cloned target gene product can be released from the GST moiety.

[0147] In an insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The SGA-56M or SGA-56Mv gene coding sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of SGA-56M or SGA-56Mv gene coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the

polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed (*e.g.*, see Smith *et al.*, 1983, *J. Virol.* 46:584; Smith, U.S. Patent No. 4,215,051).

**[0148]** In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the SGA-56M or SGA-56Mv gene coding sequence of interest may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing SGA-56M or SGA-56Mv gene product in infected hosts. (*See, e.g.*, Logan & Shenk, 1984, *Proc. Natl. Acad. Sci. USA* 81:3655). Specific initiation signals may also be required for efficient translation of inserted SGA-56M or SGA-56Mv gene product coding sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire SGA-56M or SGA-56Mv gene, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of the SGA-56M or SGA-56Mv gene coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (*See* Bittner *et al.*, 1987, *Methods in Enzymol.* 153:516).

**[0149]** In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (*e.g.*, glycosylation) and processing (*e.g.*, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins and gene products. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, and in particular, breast cancer cell lines such as, for example, BT483, Hs578T, HTB26, BT20 and T47D, and normal mammary gland cell lines such as, for example, CRL7030 and Hs578Bst.

**[0150]** For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines that stably express the SGA-56M or SGA-56Mv gene product may be engineered. Rather than using expression vectors that contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and subsequently switched

to a selective media. The selectable marker in the recombinant plasmid confers the ability to grow in selective conditions. Cells that have stably integrated the plasmid into their chromosomes grow to form foci that in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines that express the SGA-56M or SGA-56Mv gene product. Such engineered cell lines may be particularly useful in screening and identifying compounds that affect the endogenous activity of a SGA-56M and/or SGA-56Mv gene product.

**[0151]** A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler *et al.*, 1977, *Cell* 11:223), hypoxanthine-guanine phosphoribosyltransferase (Szybalska & Szybalski, 1962, *Proc. Natl. Acad. Sci. USA* 48:2026), and adenine phosphoribosyltransferase (Lowy *et al.*, 1980, *Cell* 22:817) genes can be employed in $tk^-$, $hgprt^-$ or $aprt^-$ cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for the following genes: dhfr, which confers resistance to methotrexate (Wigler *et al.*, 1980, *Proc Natl. Acad. Sci. USA* 77:3567; O'Hare *et al.*, 1981, *Proc. Natl. Acad. Sci. USA* 78:1527); gpt, which confers resistance to mycophenolic acid (Mulligan & Berg, 1981, *Proc. Natl. Acad. Sci. USA* 78:2072); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin *et al.*, 1981, *J. Mol. Biol.* 150:1); and hygro, which confers resistance to hygromycin (Santerre *et al.*, 1984, *Gene* 30:147).

### 5.2.3 SGA-56M OR SGA-56Mv TRANSGENIC ANIMALS

**[0152]** The SGA-56M or SGA-56Mv gene products can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, sheep, pigs, micro-pigs, goats, and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees may be used to generate SGA-56M or SGA-56Mv transgenic animals.

**[0153]** Transgenic animals that over- or mis-express an SGA-56M or SGA-56Mv gene product may be used in any of the methods of the invention. For example transgenic animals may be used to study the *in vivo* effects of enhanced expression levels of SGA-56M or SGA-56Mv and the onset, diagnosis or prognosis of cancer. Transgenic animals are useful for screening antagonists or agonists of SGA-56M or SGA-56Mv expression and/or activity. Transgenic animals may also be used to screen the *in vivo* effects of anti-sense or ribozyme therapeutic molecules in the treatment of cancer. Transgenic animals could be used to screen for methods of vaccinating against cancer using an SGA-56M or SGA-56Mv gene product or a portion thereof.

**[0154]** Further, SGA-56M or SGA-56Mv knock out animals are also useful in the methods of the invention. For example, animals with disruptions in only SGA-56M or both SGA-56M and SGA-56Mv can be useful in assessing the relative contribution of each of these gene products to the cancer state, as well as assessing the positive effect of a cancer therapeutic candidate.

**[0155]** For over- or mis-expression of an SGA-56M or SGA-56Mv gene product, any technique known in the art may be used to introduce the SGA-56M or SGA-56Mv gene product into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to pronuclear microinjection (Hoppe and Wagner, 1989, U.S. Pat. No. 4,873,191); retrovirus mediated gene transfer into germ lines (Van der

Putten *et al.*, 1985, *Proc. Natl. Acad. Sci. USA* 82:6148); gene targeting in embryonic stem cells (Thompson *et al.*, 1989, *Cell* 56:313); electroporation of embryos (Lo, 1983, *Mol Cell. Biol.* 3:1803); and sperm-mediated gene transfer (Lavitrano *et al.*, 1989, *Cell* 57:717); etc. For a review of such techniques, *see* Gordon, 1989, *Transgenic Animals, Intl. Rev. Cytol.* 115: 171.

**[0156]** The methods of the invention provide for the use of transgenic animals that carry the SGA-56M or SGA-56Mv transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i.e.*, mosaic animals.

**[0157]** The transgene may be integrated as a single transgene or in concatamers, *e.g.*, head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko *et al.* (Lasko *et al.*, 1992, *Proc. Natl. Acad. Sci. USA* 89:6232). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

**[0158]** When it is desired that the SGA-56M/SGA-56Mv transgene be integrated into the chromosomal site of the endogenous SGA-56M/SGA-56Mv gene, for example to disrupt the expression of SGA-56M or both SGA-56M and SGA-56Mv, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous SGA-56M/SGA-56Mv gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and partially or wholly disrupting the function of the nucleotide sequence of the endogenous SGA-56M/SGA-56Mv gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous SGA-56M/SGA-56Mv gene in only that cell type, by following, for example, the teaching of Gu *et al.* (Gu *et al.*, 1994, *Science* 265:103). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. It will be appreciated by a skilled artisan that the full length SGA-56M gene may be targeted for specific modulation using such techniques by targeting the region which is present in SGA-56M, but absent from SGA-56Mv, for homologous recombination.

**[0159]** Methods for the production of single-copy transgenic animals with chosen sites of integration are also well known to those of skill in the art. See, for example, Bronson *et al.* (Bronson, S.K. *et al.*, 1996, *Proc. Natl. Acad. Sci. USA* 93:9067).

**[0160]** Once transgenic animals have been generated, expression of the recombinant SGA-56M/SGA-56Mv gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include but are not limited to Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and RT-PCR. Samples of SGA-56M or SGA-56Mv gene-expressing tissue, may also be evaluated immunocytochemically using antibodies specific/selective for the SGA-56M or SGA-56Mv gene product.

## 5.3. ANTIBODIES TO SGA-56M OR SGA-56Mv GENE PRODUCTS

**[0161]** The methods of the present invention encompass the use of antibodies or fragments thereof capable of specifically or selectively recognizing one or more SGA-56M or SGA-56Mv gene product epitopes or epitopes of conserved variants or peptide fragments of the SGA-56M or SGA-56Mv gene products. Such antibodies may include, but are not limited to, polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, Fv fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above.

**[0162]** Such antibodies may be used, for example, in the detection of an SGA-56M or SGA-56Mv gene product in a biological sample and may, therefore, be utilized as part of a diagnostic or prognostic technique whereby patients may be tested for abnormal levels of SGA-56M or SGA-56Mv gene products, and/or for the presence of abnormal variants of the such gene products. Such antibodies may also be included as a reagent in a kit for use in a diagnostic and/or prognostic technique. Such antibodies may also be utilized in conjunction with, for example, compound screening methods, as described, below, in Section 5.5, for the evaluation of the effect of test compounds on SGA-56M or SGA-56Mv gene product levels and/or activity. Additionally, such antibodies can be used in conjunction with the gene therapy techniques described, below, in Section 5.6.4, to, for example, to evaluate the normal and/or engineered SGA-56M or SGA-56Mv-expressing cells prior to their introduction into the patient.

**[0163]** Antibodies to the SGA-56M or SGA-56Mv gene product may additionally be used in a method for the inhibition of SGA-56M or SGA-56Mv gene product activity. Thus, such antibodies may, therefore, be utilized as part of cancer treatment methods.

**[0164]** Described herein are methods for the production of antibodies or fragments thereof. Any of such antibodies or fragments thereof may be produced by standard immunological methods or by recombinant expression of nucleic acid molecules encoding the antibody or fragments thereof in an appropriate host organism.

**[0165]** For the production of antibodies against an SGA-56M or SGA-56Mv gene product, various host animals may be immunized by injection with an SGA-56M or SGA-56Mv gene product, or a portion thereof. Such host animals may include but are not limited to rabbits, mice, and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum.*

**[0166]** Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as an SGA-56M or SGA-56Mv gene product, or an antigenic functional derivative thereof. For the production of polyclonal antibodies, host animals such as those

described above, may be immunized by injection with SGA-56M or SGA-56Mv gene product supplemented with adjuvants as also described above.

**[0167]** Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, may be obtained by any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique of Kohler and Milstein, (1975, *Nature* 256:495; and U.S. Patent No. 4,376,110), the human B-cell hybridoma technique (Kosbor *et al.*, 1983, *Immunology Today* 4:72; Cole *et al.*, 1983, *Proc. Natl. Acad. Sci. USA* 80:2026), and the EBV-hybridoma technique (Cole *et al.*, 1985, *Monoclonal Antibodies And Cancer Therapy*, Alan R. Liss, Inc., pp. 77). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated *in vitro* or *in vivo*. Production of high titers of mAbs *in vivo* makes this the presently preferred method of production.

**[0168]** Techniques developed for the production of "chimeric antibodies" (Morrison et al., 1984, Proc. Natl. Acad. Sci., 81, 6851-6855; Neuberger et al., 1984, Nature 312, 604-608; Takeda et al., 1985, Nature 314, 452-454) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. (See, *e.g.*, Cabilly et al., U.S. Patent No. 4,816,567; and Boss et al., U.S. Patent No. 5,816,397). The invention thus contemplates chimeric antibodies that are specific/selective for an SGA-56M or SGA-56Mv gene product.

**[0169]** Examples of techniques that have been developed for the production of humanized antibodies are known in the art. (See, *e.g.*, Queen, U.S. Patent No. 5,585,089 and Winter, U.S. Patent No. 5,225,539) An immunoglobulin light or heavy chain variable region consists of a "framework" region interrupted by three hypervariable regions, referred to as complementarity-determining regions (CDRs). The extent of the framework region and CDRs have been precisely defined (see, "Sequences of Proteins of Immunological Interest", Kabat, E. et al., U.S. Department of Health and Human Services (1983). Briefly, humanized antibodies are antibody molecules from non-human species having one or more CDRs from the non-human species and framework regions from a human immunoglobulin molecule. The invention includes the use of humanized antibodies that are specific /selective for an SGA-56M and/or SGA-56Mv gene product in the methods of the invention.

**[0170]** The method of the invention encompasses the use of an antibody or derivative thereof comprising a heavy or light chain variable domain, said variable domain comprising (a) a set of three complementarity-determining regions (CDRs), in which said set of CDRs are from a monoclonal antibody to a gene product encoded by an SGA-56M nucleic acid sequence (SEQ ID NO: 2) or SGA-56Mv nucleic acid sequence (SEQ ID NO: 4), and (b) a set of four framework regions, in which said set of framework regions differs from the set of framework regions of the SGA-56M and/or SGA-56Mv specific monoclonal antibody, and in which said antibody or derivative thereof immunospecifically binds to the gene product encoded for by the SGA-

56M or SGA-56Mv gene sequence. Preferably, the set of framework regions is from a human monoclonal antibody, *e.g.*, a human monoclonal antibody that does not bind SGA-56M or SGA-56Mv.

**[0171]** Phage display technology can be used to increase the affinity of an antibody to an SGA-56M or SGA-56Mv gene product. This technique is useful for obtaining high affinity antibodies to an SGA-56M or SGA-56Mv gene product useful for the diagnosis and/or prognosis of a subject with cancer. The technology, referred to as affinity maturation, employs mutagenesis or CDR walking and re-selection using the SGA-56M or SGA-56Mv antigen to identify antibodies that bind with higher affinity to the antigen when compared with the initial or parental antibody (*see, e.g.,* Glaser *et al.*, 1992, *J. Immunology* 149:3903). Mutagenizing entire codons rather than single nucleotides results in a semi-randomized repertoire of amino acid mutations. Libraries can be constructed consisting of a pool of variant clones each of which differs by a single amino acid alteration in a single CDR and which contain variants representing each possible amino acid substitution for each CDR residue. Mutants with increased binding affinity for the antigen can be screened by contact with the immobilized mutants containing labeled antigen. Any screening method known in the art can be used to identify mutant antibodies with increased avidity to the antigen (*e.g.*, ELISA) (See Wu *et al.*, 1998, *Proc Natl. Acad Sci. USA* 95:6037; Yelton *et al.*, 1995, *J. Immunology* 155:1994). CDR walking may also be used to randomize the light chain (*See* Schier *et al.*, 1996, *J. Mol. Bio.* 263:551).

**[0172]** Alternatively, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778; Bird, 1988, *Science* 242:423; Huston *et al.*, 1988, *Proc. Natl. Acad. Sci. USA* 85:5879; and Ward *et al.*, 1989, *Nature* 334:544) can be adapted to produce single chain antibodies against SGA-56M or SGA-56Mv gene products. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide. Techniques for the assembly of functional Fv fragments in *E. coli* may also be used (Skerra *et al.*, 1988, *Science* 242:1038).

**[0173]** The methods of the invention include using an antibody to an SGA-56M or SGA-56Mv polypeptide, peptide or other derivative, or analog thereof that is a bispecific antibody (see generally, *e.g.*, Fanger and Drakeman, 1995, *Drug News and Perspectives* 8:133-137). Bispecific antibodies can be used for example to treat and/or prevent cancer in a subject that expresses elevated levels of an SGA-56M or SGA-56Mv gene product. Such a bispecific antibody is genetically engineered to recognize both (1) an epitope and (2) one of a variety of "trigger" molecules, *e.g.*, Fc receptors on myeloid cells, and CD3 and CD2 on T-cells, that have been identified as capable of inducing a cytotoxic T-cell to destroy a particular target. Such bispecific antibodies can be prepared either by chemical conjugation, hybridoma, or recombinant molecular biology techniques known to the skilled artisan.

**[0174]** Antibody fragments that recognize specific epitopes may be generated by known techniques. For example, such fragments include but are not limited to: $F(ab')_2$ fragments which can be produced by pepsin digestion of the antibody molecule and Fab fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed (Huse *et al.*, 1989, *Science* 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

## 5.4. USES OF THE SGA-56M AND SGA-56Mv GENE, GENE PRODUCTS, AND ANTIBODIES

**[0175]** In various embodiments, the present invention provides various uses of the SGA-56M or SGA-56Mv gene, the SGA-56M or SGA-56Mv polypeptides and peptide fragments thereof, and of antibodies directed against the SGA-56M or SGA-56Mv polypeptides and peptide fragments. Such uses include, for example, prognostic and diagnostic evaluation of cancer, and the identification of subjects with a predisposition to a cancer, as described, herein below. The invention also includes methods of treating and/or preventing cancer. The invention includes methods of vaccinating against cancer. The methods of the invention can be used for the treatment, prevention, vaccination, diagnosis, staging and/or prognosis of any cancer, or tumor, for example, but not limited to, any of the tumors or cancers listed below in Table 1.

**[0176]** Malignancies and related disorders, cells of which type can be tested *in vitro* (and/or *in vivo*), and upon observing the appropriate assay result, treated according to the methods of the present invention, include but are not limited to those listed in Table 1 (for a review of such disorders, see Fishman *et al.*, 1985, *Medicine*, 2d Ed., J.B. Lippincott Co., Philadelphia).

TABLE 1: MALIGNANCIES AND RELATED DISORDERS

- Leukemia
  - acute leukemia
    - acute lymphocytic leukemia
    - acute myelocytic leukemia
      - myeloblastic
      - promyelocytic
      - myelomonocytic
      - monocytic
      - erythroleukemia
  - chronic leukemia
    - chronic myelocytic (granulocytic) leukemia
    - chronic lymphocytic leukemia
- Polycythemia vera
- Lymphoma
  - Hodgkin's disease
  - non-Hodgkin's disease
- Multiple myeloma
- Waldenström's macroglobulinemia
- Heavy chain disease
- Solid tumors
  - sarcomas and carcinomas
    - fibrosarcoma
    - myxosarcoma
    - liposarcoma
    - chondrosarcoma
    - osteogenic sarcoma
    - chordoma
    - angiosarcoma
    - endotheliosarcoma
    - lymphangiosarcoma
    - lymphangioendotheliosarcoma
    - synovioma
    - mesothelioma
    - Ewing's tumor

TABLE 1 CONTINUED

| | |
|---|---|
| | leiomyosarcoma |
| | rhabdomyosarcoma |
| | colon carcinoma |
| | pancreatic cancer |
| | breast cancer |
| | ovarian cancer |
| | prostate cancer |
| | squamous cell carcinoma |
| | basal cell carcinoma |
| | adenocarcinoma |
| | sweat gland carcinoma |
| | sebaceous gland carcinoma |
| | papillary carcinoma |
| | papillary adenocarcinomas |
| | cystadenocarcinoma |
| | medullary carcinoma |
| | bronchogenic carcinoma |
| | renal cell carcinoma |
| | hepatoma |
| | bile duct carcinoma |
| | choriocarcinoma |
| | seminoma |
| | embryonal carcinoma |
| | Wilms' tumor |
| | cervical cancer |
| | testicular tumor |
| | lung carcinoma |
| | small cell lung carcinoma |
| | bladder carcinoma |
| | epithelial carcinoma |
| | glioma |
| | astrocytoma |
| | medulloblastoma |
| | craniopharyngioma |
| | ependymoma |
| | pinealoma |
| | hemangioblastoma |
| | acoustic neuroma |
| | oligodendroglioma |
| | menangioma |
| | melanoma |
| | neuroblastoma |
| | retinoblastoma |

---

**[0177]** In a preferred embodiment the methods of the invention are directed at diagnosis, prognosis, treatment and prevention of breast cancer. In other embodiments, the cancer is ovarian cancer, skin cancer, or cancer of the lymphoid system

**[0178]** The invention further provides for screening assays to identify antagonists or agonists of the SGA56M or SGA-56Mv gene or gene product. Thus, the invention relates to methods to identify molecules that up-regulate or down-regulate expression of the SGA-56M or SGA-56Mv gene.

**[0179]** The nucleic acid molecules, proteins, protein homologs, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) detection assays (*e.g.*, chromosomal mapping, tissue typing); c) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and d) methods of treatment (*e.g.*, therapeutic and prophylactic). For example, an SGA-56M or SGA-56Mv gene product can be used to modulate (i) cellular proliferation; (ii) cellular differentiation; and/or (iii) cellular adhesion. Isolated nucleic acid molecules that encode the SGA-56M or SGA-56Mv gene or a fragment or an open reading frame thereof can be used to express proteins (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect mRNA (*e.g.*, in a biological sample) or a genetic lesion, and to modulate activity of an SGA-56M or SGA-56Mv polypeptide. In addition, an SGA-56M or SGA-56Mv gene product can be used to screen drugs or compounds which modulate activity or expression of the SGA-56M or SGA-56Mv gene product as well as to treat disorders characterized by insufficient or excessive production of the SGA-56M or SGA-56Mv gene product or production of a form the SGA-56M or SGA-56Mv gene product which has decreased or aberrant activity compared to the wild type protein. In addition, the antibodies that specifically or selectively bind to an SGA-56M or SGA-56Mv gene product can be used to detect, isolate, and modulate activity of the SGA-56M or SGA-56Mv gene product.

**[0180]** In one embodiment, the present invention provides a variety of methods for the diagnostic and prognostic evaluation of cancer, including breast cancer. Such methods may, for example, utilize reagents such as the SGA-56M or SGA-56Mv gene nucleotide sequences described in Sections 5.1, and antibodies directed against SGA-56M or SGA-56Mv gene products, including peptide fragments thereof, as described, above, in Section 5.2. Specifically, such reagents may be used, for example, for: (1) the detection of the presence of SGA-56M or SGA-56Mv gene mutations, or the detection of either over- or under-expression of SGA-56M or SGA-56Mv gene mRNA, preneoplastic or neoplastic, relative to normal cells or the qualitative or quantitative detection of other allelic forms of SGA-56M or SGA-56Mv transcripts which may correlate with breast cancer or susceptibility toward neoplastic changes, and (2) the detection of an over-abundance of an SGA-56M or SGA-56Mv gene product relative to a non-diseased state or relative to a predetermined non-cancerous standard or the presence of a modified (*e.g.*, less than full-length) SGA-56M or SGA-56Mv gene product which correlates with a neoplastic state or a progression toward neoplasia or metastasis.

**[0181]** The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic test kits comprising at least one specific or selective SGA-56M or SGA-56Mv gene nucleic acid or anti-SGA-56M or anti-SGA-56Mv antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings or in home settings, to diagnose patients exhibiting preneoplastic or neoplastic abnormalities, and to screen and identify those individuals exhibiting a predisposition to such neoplastic changes.

**[0182]** Nucleic acid-based detection techniques are described, below, in Section 5.4.1. Peptide detection techniques are described, below, in Section 5.4.2.

### 5.4.1. DETECTION OF SGA-56M OR SGA-56Mv GENE NUCLEIC ACID MOLECULES

**[0183]** In a preferred embodiment, the invention involves methods to assess quantitative and qualitative aspects of SGA-56M or SGA-56Mv gene expression. In one example the increased expression of an SGA-56M or SGA-56Mv gene or gene product indicates a predisposition for the development of cancer. Alternatively, enhanced expression levels of an SGA-56M or SGA-56Mv gene or gene product can indicate the presence of cancer in a subject or the risk of metastasis of said cancer in said subject. Techniques well known in the art, *e.g.*, quantitative or semi-quantitative RT PCR or Northern blot, can be used to measure expression levels of SGA-56M or SGA-56Mv. Methods that describe both qualitative and quantitative aspects of SGA-56M or SGA-56Mv gene or gene product expression are described in detail in the examples *infra*. The measurement of SGA-56M or SGA-56Mv gene expression levels can include measuring naturally occurring SGA-56M or SGA-56Mv transcripts and variants thereof as well as non-naturally occurring variants thereof, however for the diagnosis and/or prognosis of cancer in a subject the SGA-56M or SGA-56Mv gene product is preferably a naturally occurring SGA-56M or SGA-56Mv gene product or variant thereof. Thus, the invention relates to methods of diagnosing and/or predicting cancer in a subject by measuring the expression of the SGA-56M or SGA-56Mv gene in a subject. For example an increased level of mRNA encoded by a SGA-56M or SGA-56Mv nucleic acid sequence (e.g., SEQ ID NO: 1 or SEQ ID NO: 3), or other gene product, as compared to a non-cancerous sample or a non-cancerous predetermined standard would indicate the presence of cancer in said subject or the increased risk of developing cancer in said subject.

**[0184]** In another example the increased level of mRNA encoded for by an SGA-56M or SGA-56Mv nucleic acid sequence (e.g., SEQ ID NO: 1 or SEQ ID NO: 3), or other gene product, as compared to a non-cancerous sample or a non-cancerous predetermined standard would indicate the risk of metastasis in a cancer subject or the likelihood of a poor prognosis in said subject.

**[0185]** In another example, RNA from a cell type or tissue known, or suspected, to express the SGA-56M or SGA-56Mv gene, such as breast cancer cells, or other types of cancer cells, including metastases, may be isolated and tested utilizing hybridization or PCR techniques as described, above. The isolated cells can be derived from cell culture or from a patient. The analysis of cells taken from culture may be a necessary step in the assessment of cells to be used as part of a cell-based gene therapy technique or, alternatively, to test the effect of compounds on the expression of the SGA-56M or SGA-56Mv gene. Such analyses may reveal both quantitative and qualitative aspects of the expression pattern of the SGA-56M or SGA-56Mv gene, including activation or inactivation of SGA-56M or SGA-56Mv gene expression and presence of alternatively spliced SGA-56M or SGA-56Mv transcripts.

**[0186]** In one embodiment of such a detection scheme, a cDNA molecule is synthesized from an RNA molecule of interest by reverse transcription. All or part of the resulting cDNA is then used as a template for a nucleic acid amplification reaction, such as a PCR or the like. The nucleic acid reagents used as synthesis initiation reagents (*e.g.*, primers) in the reverse transcription and nucleic acid amplification steps of this method are chosen from among the SGA-56M or SGA-56Mv gene nucleic acid reagents described in Section 5.1. The preferred lengths of such nucleic acid reagents are at least 9-30 nucleotides.

**[0187]** For detection of the amplified product, the nucleic acid amplification may be performed using radioactively or non-radioactively labeled nucleotides. Alternatively, enough amplified product may be made such that the product may be visualized by standard ethidium bromide staining or by utilizing any other suitable nucleic acid staining method.

**[0188]** RT-PCR techniques can be utilized to detect differences in SGA-56M or SGA-56Mv transcript size that may be due to normal or abnormal alternative splicing. Additionally, such techniques can be performed using standard techniques to detect quantitative differences between levels of SGA-56M or SGA-56Mv transcripts detected in normal individuals relative to those individuals having cancer or exhibiting a predisposition toward neoplastic changes.

**[0189]** In the case where detection of particular alternatively spliced species is desired, appropriate primers and/or hybridization probes can be used, such that, in the absence of such a sequence, for example, no amplification products are generated. Alternatively, primer pairs may be chosen utilizing the sequence data depicted in FIG. 1 or FIG. 2 to yield fragments of differing size depending on whether a particular exon is present or absent from the transcript of SGA-56M or SGA-56Mv being analyzed.

**[0190]** As an alternative to amplification techniques, standard Northern analyses can be performed if a sufficient quantity of the appropriate cells can be obtained. The preferred length of a probe used in a Northern analysis may, for example, be 9-50 nucleotides. Utilizing such techniques, quantitative as well as size related differences between SGA-56M or SGA-56Mv transcripts can also be detected.

**[0191]** Additionally, it is possible to perform such SGA-56M or SGA-56Mv gene expression assays *in situ*, *i.e.*, directly upon tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections, such that no nucleic acid purification is necessary. Nucleic acid reagents such as those described in Section 5.1 may be used as probes and/or primers for such *in situ* procedures (*see, e.g.*, Nuovo, G.J., 1992, *PCR In Situ Hybridization: Protocols And Applications*, Raven Press, NY).

**[0192]** Mutations or polymorphisms within the SGA-56M or SGA-56Mv gene can be detected by utilizing a number of techniques. Nucleic acid from any nucleated cell can be used as the starting point for such assay techniques, and may be isolated according to standard nucleic acid preparation procedures that are well known to those of skill in the art. For the detection of SGA-56M or SGA-56Mv mutations, any nucleated cell can be used as a starting source for genomic nucleic acid. For the detection of SGA-56M or SGA-56Mv transcripts or SGA-56M or SGA-56Mv gene products, any cell type or tissue in which the SGA-56M or SGA-56Mv gene is expressed, such as, for example, breast cancer cells, including metastases, may be utilized.

**[0193]** Genomic DNA may be used in hybridization or amplification assays of biological samples to detect abnormalities involving SGA-56M or SGA-56Mv gene structure, including point mutations, insertions, deletions and chromosomal rearrangements. Such assays may include, but are not limited to, direct sequencing (Wong, C. *et al.*, 1987, *Nature* 330:384), single stranded conformational polymorphism analyses (SSCP; Orita, M. *et al.*, 1989, *Proc. Natl. Acad. Sci. USA* 86:2766), heteroduplex analysis (Keen, T.J. *et al.*, 1991, *Genomics* 11:199; Perry, D.J. & Carrell, R.W., 1992), denaturing gradient gel electrophoresis (DGGE; Myers, R.M. *et al.*, 1985, *Nucl. Acids Res.* 13:3131), chemical mismatch cleavage (Cotton, R.G. *et al.*, 1988,

*Proc. Natl. Acad. Sci. USA* 85:4397) and oligonucleotide hybridization (Wallace, R.B. *et al.*, 1981, *Nucl. Acids Res.* 9:879; Lipshutz, R.J. *et al.*, 1995, *Biotechniques* 19:442).

**[0194]** Diagnostic methods for the detection of SGA-56M or SGA-56Mv nucleic acid molecules, in patient samples or other appropriate cell sources, may involve the amplification of specific gene sequences, *e.g.*, by the polymerase chain reaction (PCR; *See* Mullis, K.B., 1987, U.S. Patent No. 4,683,202), followed by the analysis of the amplified molecules using techniques well known to those of skill in the art, such as, for example, those listed above. Utilizing analysis techniques such as these, the amplified sequences can be compared to those that would be expected if the nucleic acid being amplified contained only normal copies of the SGA-56M or SGA-56Mv gene in order to determine whether an SGA-56M or SGA-56Mv gene mutation exists.

**[0195]** Further, well-known genotyping techniques can be performed to type polymorphisms that are in close proximity to mutations in the SGA-56M or SGA-56Mv gene itself. These polymorphisms can be used to identify individuals in families likely to carry mutations. If a polymorphism exhibits linkage disequilibrium with mutations in the SGA-56M or SGA-56Mv gene, it can also be used to identify individuals in the general population likely to carry mutations. Polymorphisms that can be used in this way include restriction fragment length polymorphisms (RFLPs), which involve sequence variations in restriction enzyme target sequences, single-nucleotide polymorphisms (SNPs) and simple sequence repeat polymorphisms (SSLPs).

**[0196]** For example, Weber (U.S. Pat. No. 5,075,217) describes a DNA marker based on length polymorphisms in blocks of (dC-dA)n-(dG-dT)n short tandem repeats. The average separation of (dC-dA)n-(dG-dT)n blocks is estimated to be 30,000-60,000 bp. Markers that are so closely spaced exhibit a high frequency of co-inheritance, and are extremely useful in the identification of genetic mutations, such as, for example, mutations within the SGA-56M or SGA-56Mv gene, and the diagnosis of diseases and disorders related to SGA-56M or SGA-56Mv mutations.

**[0197]** Also, Caskey *et al.* (U.S. Pat.No. 5,364,759), describe a DNA profiling assay for detecting short tri- and tetra-nucleotide repeat sequences. The process includes extracting the DNA of interest, such as the SGA-56M or SGA-56Mv gene, amplifying the extracted DNA, and labeling the repeat sequences to form a genotypic map of the individual's DNA.

**[0198]** An SGA-56M or SGA-56Mv probe could be used to directly identify RFLPs. Additionally, an SGA-56M or SGA-56Mv probe or primers derived from the SGA-56M or SGA-56Mv sequence could be used to isolate genomic clones such as YACs, BACs, PACs, cosmids, phage or plasmids. The DNA contained in these clones can be screened for SNPs or SSLPs using standard hybridization or sequencing procedures.

**[0199]** Alternative diagnostic methods for the detection of SGA-56M or SGA-56Mv gene expression, SGA-56M or SGA-56Mv gene mutations or polymorphisms can include hybridization techniques which involve for example, contacting and incubating nucleic acids including recombinant DNA molecules, cloned genes or degenerate variants thereof, obtained from a sample, *e.g.*, derived from a patient sample or other appropriate cellular source, with one or more labeled nucleic acid reagents including recombinant DNA molecules, cloned genes or degenerate variants thereof, as described in Section 5.1, under conditions favorable for the specific or

selective annealing of these reagents to their complementary sequences within the SGA-56M or SGA-56Mv gene. Preferably, the lengths of these nucleic acid reagents are at least 9 to 50 nucleotides. After incubation, all non-annealed nucleic acids are removed from the nucleic acid: SGA-56M or SGA-56Mv molecule hybrid. The presence of nucleic acids that have hybridized, if any such molecules exist, is then detected. Using such a detection scheme, a nucleic acid from the cell type or tissue of interest can be immobilized, for example, to a solid support such as a membrane, or a plastic surface such as that on a microtiter plate or polystyrene beads or to a glass surface such as a microscope slide. In this case, after incubation, non-annealed, labeled nucleic acid reagents of the type described in Section 5.1 are easily removed. Detection of the remaining, annealed, labeled SGA-56M or SGA-56Mv nucleic acid reagents is accomplished using standard techniques well known in the art. The SGA-56M or SGA-56Mv gene sequences to which the nucleic acid reagents have annealed can be compared to the annealing pattern expected from a normal SGA-56M or SGA-56Mv gene sequence in order to determine whether an SGA-56M or SGA-56Mv gene mutation is present.

### 5.4.2. DETECTION OF SGA-56M AND SGA-56Mv ENCODED PROTEINS

**[0200]** Detection of the SGA-56M or SGA-56Mv gene product includes the detection of the proteins comprising SEQ ID NO: 5 or SEQ ID NO: 6. Detection of elevated levels of SGA-56M or SGA-56Mv, compared to a non-cancerous sample or a non-cancerous predetermined standard can indicate the presence of cancer, or predisposition to developing cancer in a subject. Detection of elevated levels of said protein in a subject compared to a non-cancerous sample or a non-cancerous predetermined standard can indicate the likelihood of metastasis of a cancer in the subject, and/or poor prognosis for the subject. The diagnosis and/or prognosis of cancer pertains to the detection of naturally occurring SGA-56M or SGA-56Mv polypeptides in a subject. Detection of an SGA-56M or SGA-56Mv polypeptide can be by any method known in the art.

**[0201]** Antibodies directed against naturally occurring SGA-56M or SGA-56Mv, or naturally occurring variants thereof or peptide fragments thereof, which are discussed, above, in Section 5.2, may be used as diagnostics and prognostics, as described herein. Such diagnostic methods, may be used to detect abnormalities in the level of SGA-56M or SGA-56Mv gene expression, or abnormalities in the structure and/or temporal, tissue, cellular, or subcellular location of the SGA-56M or SGA-56Mv-encoded polypeptide. Antibodies, or fragments of antibodies, such as those described herein, may be used to screen potentially therapeutic compounds *in vitro* to determine their effects on SGA-56M or SGA-56Mv gene expression and SGA-56M or SGA-56Mv-encoded polypeptide production. The compounds that have beneficial effects on cancer, *e.g.*, breast cancer can be identified and a therapeutically effective dose determined.

**[0202]** The tissue or cell type to be analyzed will generally include those which are known, or suspected, to express the SGA-56M or SGA-56Mv gene, such as, for example, cancer cells including breast cancer cells, ovarian cancer cells, skin cancer cells, lymphoid cancer cells, and metastatic forms thereof. The protein isolation methods employed herein may, for example, be such as those described in Harlow and Lane (Harlow, E. and Lane, D., 1988, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York). The isolated cells can be derived from cell culture or from a patient. The

analysis of cells taken from culture may be a necessary step to test the effect of compounds on the expression of the SGA-56M or SGA-56Mv gene.

**[0203]** Preferred diagnostic methods for the detection of SGA-56M or SGA-56Mv gene products or conserved variants or peptide fragments thereof, may involve, for example, immunoassays wherein the SGA-56M or SGA-56Mv gene products or conserved variants, including gene products which are the result of alternatively spliced transcripts, or peptide fragments are detected by their interaction with an anti-SGA-56M or anti-SGA-56Mv gene product specific antibody.

**[0204]** For example, antibodies, or fragments of antibodies, such as those described above in Section 5.3, useful in the present invention may be used to quantitatively or qualitatively detect the presence of SGA-56M or SGA-56Mv-encoded polypeptides or naturally occurring variants or peptide fragments thereof. The antibodies (or fragments thereof) useful in the present invention may, additionally, be employed histologically, as in immunofluorescence or immunoelectron microscopy, for *in situ* detection of SGA-56M or SGA-56Mv gene products or conserved variants or peptide fragments thereof. *In situ* detection may be accomplished by removing a histological specimen from a subject, such as paraffin embedded sections of tissue, *e.g.*, breast tissues, and applying thereto a labeled antibody of the present invention. The antibody (or fragment) is preferably applied by overlaying the labeled antibody (or fragment) onto a biological sample. Since the SGA-56M or SGA-56Mv gene product is present in the cytoplasm, it may be desirable to introduce the antibody inside the cell, for example, by making the cell membrane permeable. The SGA-56M or SGA-56Mv polypeptides may also be expressed on the cell surface, thus cells can be directly labeled by applying antibodies that are specific or selective for the SGA-56M or SGA-56Mv polypeptides or fragment thereof to the cell surface.

**[0205]** Through the use of such a procedure, it is possible to determine not only the presence of the SGA-56M or SGA-56Mv gene product, or naturally occurring variants thereof or peptide fragments, but also its distribution in the examined tissue. Using the methods of the present invention, those of ordinary skill will readily perceive that any of a wide variety of histological methods (such as staining procedures) can be modified in order to achieve such *in situ* detection.

**[0206]** Immunoassays for SGA-56M or SGA-56Mv-encoded polypeptides or conserved variants or peptide fragments thereof will typically comprise contacting a sample, such as a biological fluid, tissue or a tissue extract, freshly harvested cells, or lysates of cells which have been incubated in cell culture, in the presence of an antibody that specifically or selectively binds to an SGA-56M or SGA-56Mv gene product, *e.g.*, a detectably labeled antibody capable of identifying SGA-56M or SGA-56Mv polypeptides or conserved variants or peptide fragments thereof, and detecting the bound antibody by any of a number of techniques well-known in the art (*e.g.*, Western blot, ELISA, FACS).

**[0207]** The biological sample may be brought in contact with and immobilized onto a solid phase support or carrier such as nitrocellulose, or other solid support that is capable of immobilizing cells, cell particles or soluble proteins. The support may then be washed with suitable buffers followed by treatment with the detectably labeled antibody that selectively or specifically binds to an SGA-56M or SGA-56Mv-encoded

polypeptide. The solid phase support may then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on solid support may then be detected by conventional means.

**[0208]** By "solid phase support or carrier" is intended any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention. The support material may have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, *etc.* Preferred supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding antibody or antigen, or will be able to ascertain the same by use of routine experimentation.

**[0209]** The anti-SGA-56M or anti-SGA-56Mv antibody can be detectably labeled by linking the same to an enzyme and using the labeled antibody in an enzyme immunoassay (EIA) (Voller, A., "The Enzyme Linked Immunosorbent Assay (ELISA)", 1978, *Diagnostic Horizons* 2:1, Microbiological Associates Quarterly Publication, Walkersville, MD); Voller, A. *et al.*, 1978, *J. Clin. Pathol.* 31: 507-520; Butler, J.E., 1981, *Meth. Enzymol.* 73:482; Maggio, E. (ed.), 1980, *Enzyme Immunoassay*, CRC Press, Boca Raton, FL,; Ishikawa, E. *et al.*, (eds.), 1981, *Enzyme Immunoassay*, Kgaku Shoin, Tokyo). The enzyme that is bound to the antibody will react with an appropriate substrate, preferably a chromogenic substrate, in such a manner as to produce a chemical moiety that can be detected, for example, by spectrophotometric or fluorimetric means. Enzymes which can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycerophosphate, dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods that employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

**[0210]** Detection may also be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, it is possible to detect SGA-56M or SGA-56Mv-encoded polyepeptides through the use of a radioimmunoassay (RIA) (see, for example, Weintraub, B., *Principles of Radioimmunoassays*, *Seventh Training Course on Radioligand Assay Techniques*, The Endocrine Society, March, 1986). The radioactive isotope can be detected by such means as the use of a gamma counter, a scintillation counter, or by autoradiography.

**[0211]** It is also possible to label the antibody with a fluorescent compound. When the fluorescently labeled antibody is exposed to light of the proper wavelength, its presence can then be detected due to fluorescence emission. Among the most commonly used fluorescent labeling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, *o*-phthaldehyde and fluorescamine.

[0212] The antibody can also be detectably labeled using fluorescence emitting metals such as $^{152}Eu$, or others of the lanthanide series. These metals can be attached to the antibody using such metal chelating groups as diethylenetriaminepentacetic acid (DTPA) or ethylenediaminetetraacetic acid (EDTA).

[0213] The antibody can also be detectably labeled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of particularly useful chemiluminescent labeling compounds are luminol, isoluminol, theromatic acridinium ester, imidazole, acridinium salt and oxalate ester.

[0214] Likewise, a bioluminescent compound may be used to label the antibody of the present invention. Bioluminescence is a type of chemiluminescence found in biological systems wherein a catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin.

[0215] In various embodiments, the present invention provides methods for the measurement of SGA-56M or SGA-56Mv polypeptides, and the uses of such measurements in clinical applications using SGA-56M or SGA-56Mv specific antibodies.

[0216] The measurement of SGA-56M or SGA-56Mv polyppeptides of the invention can be valuable in detecting and/or staging breast cancer and other cancers in a subject, in screening of breast cancer and other cancers in a population, in differential diagnosis of the physiological condition of a subject, and in monitoring the effect of a therapeutic treatment on a subject.

[0217] The present invention also provides for detecting, diagnosing, or staging of breast cancer and other cancers, or the monitoring of treatment of breast cancer and other cancers by measuring the level of expression of an SGA-56M or SGA-56Mv polypeptide. In addition to the SGA-56M or SGA-56Mv polypeptide at least one other marker, such as a receptor or differentiation antigen can also be measured. For example, serum markers selected from, for example but not limited to, carcinoembryonic antigen (CEA), CA15-3, CA549, CAM26, M29, CA27.29 and MCA can be measured in combination with an SGA-56M or SGA-56Mv polypeptide to detect, diagnose, stage, and/or monitor treatment of breast cancer and other cancers. In another embodiment, the prognostic indicator is the observed change in different marker levels relative to one another, rather than the absolute levels of the markers present at any one time. These measurements can also aid in predicting therapeutic outcome and in evaluating and monitoring the overall disease status of a subject.

[0218] In a specific embodiment of the invention, soluble SGA-56M or SGA-56Mv polypeptide alone or in combination with other markers can be measured in any body fluid of the subject including but not limited to blood, serum, plasma, milk, urine, saliva, pleural effusions, synovial fluid, spinal fluid, tissue infiltrations and tumor infiltrates. In another embodiment an SGA-56M or SGA-56Mv polypeptide is measured in tissue samples or cells directly. The present invention also contemplates a kit for measuring the level of SGA-56M or SGA-56Mv expression in a biological sample and the use of said kit to diagnose a subject with cancer.

Alternatively said kit could be used to determine the prognosis of a cancer patient or the risk of metastasis of said cancer.

**[0219]** Any of numerous immunoassays can be used in the practice of the methods of the instant invention, such as those described in Section 5.4.2. Antibodies, or antibody fragments containing the binding domain, which can be employed include, but are not limited to, suitable antibodies among those in Section 5.3 and other antibodies known in the art or which can be obtained by procedures standard in the art such as those described in Section 5.3.

### 5.4.2.1 IN VIVO IMAGING USING ANTIBODIES TO AN SGA-56M OR SGA-56Mv POLYPEPTIDE

**[0220]** Current diagnostic and therapeutic methods make use of antibodies to target imaging agents or therapeutic substances, *e.g.*, to tumors. Thus, labeled antibodies immunologically specific for an SGA-56M or SGA-56Mv polypepeptide can be used in the methods of the invention for the *in vivo* imaging, detection, and treatment of cancer in a subject.

**[0221]** Antibodies may be linked to chelators such as those described in U.S. Patent No. 4,741,900 or U.S. Patent No. 5,326,856. The antibody-chelator complex may then be radiolabeled to provide an imaging agent for diagnosis and/or treatment of disease. The antibodies may also be used in the methods that are disclosed in U.S. Patent No. 5,449,761 for creating a radiolabeled antibody for use in imaging or radiotherapy.

**[0222]** In *in vivo* diagnostic applications, specific tissues or even specific cellular disorders, *e.g.*, cancer, may be imaged by administration of a sufficient amount of a labeled antibody using the methods of the instant invention.

**[0223]** A wide variety of metal ions suitable for *in vivo* tissue imaging have been tested and utilized clinically. For imaging with radioisotopes, the following characteristics are generally desirable: (a) low radiation dose to the patient; (b) high photon yield which permits a nuclear medicine procedure to be performed in a short time period; (c) ability to be produced in sufficient quantities; (d) acceptable cost; (e) simple preparation for administration; and (f) no requirement that the patient be sequestered subsequently. These characteristics generally translate into the following: (a) the radiation exposure to the most critical organ is less than 5 rad; (b) a single image can be obtained within several hours after infusion; (c) the radioisotope does not decay by emission of a particle; (d) the isotope can be readily detected; and (e) the half-life is less than four days (Lamb and Kramer, "Commercial Production of Radioisotopes for Nuclear Medicine", In Radiotracers For Medical Applications, Vol. 1, Rayudu (Ed.), CRC Press, Inc., Boca Raton, pp. 17-62). Preferably, the metal is technetium-99m.

**[0224]** By way of illustration, the targets that one may image include any solid neoplasm, certain organs such as lymph nodes, parathyroids, spleen and kidney, sites of inflammation or infection (*e.g.*, macrophages at such sites), myocardial infarction or thromboses (neoantigenic determinants on fibrin or platelets), and the like evident to one of ordinary skill in the art. Furthermore, the neoplastic tissue may be present in bone, internal organs, connective tissue, or skin.

[0225] As is also apparent to one of ordinary skill in the art, one may use the methods of the present invention in *in vivo* therapeutics (*e.g.*, using radiotherapeutic metal complexes), especially after having diagnosed a diseased condition via the *in vivo* diagnostic method described above, or in *in vitro* diagnostic application (*e.g.*, using a radiometal or a fluorescent metal complex).

[0226] Accordingly, a method of diagnosing cancer by obtaining an image of an internal region of a subject is contemplated in the instant invention which comprises administering to a subject an effective amount of an antibody composition specific for an SGA-56M or SGA-56Mv polypeptide conjugated with a metal in which the metal is radioactive, and recording the scintigraphic image obtained from the decay of the radioactive metal. Likewise, a method is contemplated of enhancing a magnetic resonance image (MRI) of an internal region of a subject which comprises administering to a subject an effective amount of an antibody composition containing a metal in which the metal is paramagnetic, and recording the MRI of an internal region of the subject.

[0227] Other methods are directed to enhancing a sonographic image of an internal region of a subject comprising administering to a subject an effective amount of an antibody composition containing a metal and recording the sonographic image of an internal region of the subject. In this latter application, the metal is preferably any non-toxic heavy metal ion. A method of enhancing an X-ray image of an internal region of a subject is also provided which comprises administering to a subject an antibody composition containing a metal, and recording the X-ray image of an internal region of the subject. A radioactive, non-toxic heavy metal ion is preferred.

### 5.4.3. DETECTING AND STAGING CANCER IN A SUBJECT

[0228] The methods of the present invention include measurement of a naturally occurring SGA-56M or SGA-56Mv polypeptide, or naturally occurring variants thereof, or fragment thereof, soluble SGA-56M or SGA-56Mv polypeptide or intracellular SGA-56M or SGA-56Mv polypeptides to detect breast cancer or other cancers in a subject or to stage breast cancer or other cancers in a subject.

[0229] Staging refers to the grouping of patients according to the extent of their disease. Staging is useful in choosing treatment for individual patients, estimating prognosis, and comparing the results of different treatment programs. Staging of breast cancer for example is performed initially on a clinical basis, according to the physical examination and laboratory radiologic evaluation. The most widely used clinical staging system is the one adopted by the International Union against Cancer (UICC) and the American Joint Committee on Cancer (AJCC) Staging and End Results Reporting. It is based on the tumor-nodes-metastases (TNM) system as detailed in the 1988 *Manual for Staging of Cancer.* Breast cancer diseases or conditions that may be detected and/or staged in a subject according to the present invention include but are not limited to those listed in Table 2.

TABLE 2

STAGING OF BREAST CANCER

| | |
|---|---|
| T | PRIMARY TUMORS |
| TX | Primary tumor cannot be assessed |
| T0 | No evidence of primary tumor |
| Tis | Carcinoma in situ: intraductal carcinoma, lobular carcinoma, or Paget's disease with no tumor |
| T1 | Tumor 2 cm or less in its greatest dimension<br>a. 0.5 cm or less in greatest dimension<br>b. Larger than 0.5 cm, but not larger than 1 cm in greatest dimension<br>c. Larger than 1 cm, but not larger than 2 cm in greatest dimension |
| T2 | Tumor more than 2 cm but not more than 5 cm in greatest dimension |
| T3 | Tumor more than 5cm in its greatest dimension |
| T4 | Tumor of any size with direct extension to chest wall or to skin. Chest wall includes ribs, intercostal muscles, and serratus anterior muscle, but not pectoral muscle.<br>a. Extension to chest wall<br>b. Edema (including peau d'orange), ulceration of the skin of the breast, or satellite skin nodules confined to the same breast<br>c. Both of the above<br>d. Inflammatory carcinoma |

Dimpling of the skin, nipple retraction, or any other skin changes except those in T4b may occur in T1, T2 or T3 without affecting the classification.

| | |
|---|---|
| N | REGIONAL LYMPH NODES |
| NX | Regional lymph nodes cannot be assessed (*e.g.*, previously removed) |
| N0 | No regional lymph node metastases |
| N1 | Metastasis to movable ipsilateral axillary node(s) |
| N2 | Metastases to ipsilateral axillary nodes fixed to one another or to other structures |
| N3 | Metastases to ipsilateral internal mammary lymph node(s) |
| M | DISTANT METASTASIS |
| M0 | No evidence of distant metastasis |
| M1 | Distant metastases (including metastases to ipsilateral supraclavicular lymph nodes) |

---

**[0230]** Any immunoassay, such as those described in Section 5.4.2 can be used to measure the amount of SGA-56M or SGA-56Mv polypeptide or soluble SGA-56M or SGA-56Mv polypeptide as compared to a baseline level. This baseline level can be the amount that is established to be normally present in the tissue or body fluid of subjects with various degrees of the disease or disorder. An amount present in the tissue or body fluid of the subject that is similar to a standard amount, established to be normally present in the tissue or body fluid of the subject during a specific stage of cancer or breast cancer, is indicative of the stage of the

disease in the subject. The baseline level could also be the level present in the subject prior to the onset of disease or the amount present during remission of the disease.

**[0231]** In specific embodiments of this aspect of the invention, measurements of levels of the SGA-56M or SGA-56Mv polypeptide or soluble SGA-56M or SGA-56Mv polypeptide can be used in the detection of infiltrative ductal carcinoma (IDC), the presence of metastases, or both. Increased levels of SGA-56M or SGA-56Mv polypeptides or soluble SGA-56M or SGA-56Mv polypeptide are associated with metastases.

**[0232]** In another embodiment of the invention, the measurement of soluble SGA-56M or SGA-56Mv polypeptide, intra-cellular SGA-56M or SGA-56Mv polypeptide, fragments thereof or immunologically related molecules can be used to differentially diagnose in a subject a particular disease phenotype or physiological condition as distinct as from among two or more phenotypes or physiological conditions. For example, measurements of SGA-56M or SGA-56Mv polypeptide or soluble SGA-56M or SGA-56Mv polypeptide levels may be used in the differential diagnosis of infiltrative ductal carcinoma, as distinguished from ductal carcinoma in situ or benign fibroadenomas. To this end, for example, the measured amount of the SGA-56M or SGA-56Mv polypeptide is compared with the amount of the molecule normally present in the tissue, cells or body fluid of a subject with one of the suspected physiological conditions. A measured amount of the SGA-56M or SGA-56Mv polypeptide similar to the amount normally present in a subject with one of the physiological conditions, and not normally present in a subject with one or more of the other physiological conditions, is indicative of the physiological condition of the subject.

**[0233]** As an alternative to measuring levels of SGA-56M or SGA-56Mv polypeptides in the foregoing staging methods, levels of SGA-56M or SGA-56Mv transcript can be measured, for example by the methods described in Section 5.4.1, *supra*.

### 5.4.4. MONITORING THE EFFECT OF A THERAPEUTIC TREATMENT

**[0234]** The present invention provides a method for monitoring the effect of a therapeutic treatment on a subject who has undergone the therapeutic treatment.

**[0235]** Clinicians very much need a procedure that can be used to monitor the efficacy of cancer treatments. SGA-56M or SGA-56Mv-encoded polypeptides and/or transcripts can be identified and detected in breast cancer patients or other cancer patients with different manifestations of disease, providing a sensitive assay to monitor therapy. The therapeutic treatments which may be evaluated according to the present invention include but are not limited to radiotherapy, surgery, chemotherapy, vaccine administration, endocrine therapy, immunotherapy, and gene therapy, *etc.* The chemotherapeutic regimens include, but are not limited to administration of drugs such as, for example, methotrexate, fluorouracil, cyclophosphamide, doxorubicin, and taxol. The endocrine therapeutic regimens include, but are not limited to administration of tamoxifen, progestins, etc.

**[0236]** The method of the invention comprises measuring at suitable time intervals before, during, or after therapy, the amount of an SGA-56M or SGA-56Mv transcript or polypeptide (including soluble polypeptide), or any combination of the foregoing. Any change or absence of change in the absolute or relative amounts of

the SGA-56M or SGA-56Mv gene products can be identified and correlated with the effect of the treatment on the subject.

**[0237]** In particular, the serum- or cell-associated levels of an SGA-56M or SGA-56Mv-encoded polypeptide relates to the severity of a cancer, such as breast cancer, risk of metastasis of said cancer, and poor prognosis. Since serum- or cell-associated SGA-56M or SGA-56Mv polypeptide levels are generally undetectable or negligible in normal individuals, generally, a decrease in the level of detectable SGA-56M or SGA-56Mv polypeptide after a therapeutic treatment is associated with efficacious treatment.

**[0238]** In a preferred aspect, the approach that can be taken is to determine the levels of soluble or cell associated SGA-56M or SGA-56Mv polyepeptide levels at different time points and to compare these values with a baseline level. The baseline level can be either the level of the SGA-56M or SGA-56Mv polypeptide present in normal, disease free individuals; and/or the levels present prior to treatment, or during remission of disease, or during periods of stability. These levels can then be correlated with the disease course or treatment outcome.

### 5.4.5. PROGNOSTIC ASSAYS

**[0239]** The methods described herein can furthermore be utilized as prognostic assays to identify subjects having or at risk of developing cancer or another disease or disorder associated with aberrant expression or activity of an SGA-56M or SGA-56Mv polypeptide. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing cancer, *e.g.*, breast cancer, or another disorder associated with aberrant expression or activity of an SGA-56M or SGA-56Mv polypeptide. Thus, the present invention provides a method in which a test sample is obtained from a subject and an SGA-56M or SGA-56Mv polypeptide or nucleic acid (*e.g.*, mRNA, genomic DNA) of the invention is detected, wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the SGA-56M or SGA-56Mv polypeptide, *e.g.*, cancer. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

**[0240]** The prognostic assays described herein, for example, can be used to identify a subject having or at risk of developing disorders such as cancers, for example, hormone-sensitive cancer such as breast cancer.

**[0241]** In another example, prognostic assays described herein can be used to identify a subject having or at risk of developing related disorders associated with expression of polypeptides or nucleic acids of the invention.

**[0242]** Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat cancer or another disease or disorder associated with aberrant expression or activity of an SGA-56M or SGA-56Mv polypeptide. For example, such methods can be used to determine whether a subject can be effectively treated with a specific agent or class of agents (*e.g.*, agents of a

type which decrease activity or expression level of an SGA-56M or SGA-56Mv transcript or polypeptide). Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant expression or activity of the SGA-56M or SGA-56Mv transcript or polypeptide in which a test sample is obtained and the polypeptide or nucleic acid encoding the polypeptide is detected (*e.g.*, wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant expression or activity of the SGA-56M or SGA-56Mv transcript or polypeptide).

**[0243]** The methods of the invention can also be used to detect genetic lesions or mutations in an SGA-56M or SGA-56Mv gene, thereby determining if a subject with the lesioned gene is at increased or reduced risk for a disorder characterized by aberrant expression or activity of a polypeptide of the invention, *e.g.*, cancer. In one embodiment, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion or mutation characterized by at least one of an alteration affecting the integrity of a gene encoding an SGA-56M or SGA-56Mv polypeptide, or the mis-expression of the gene encoding an SGA-56M or SGA-56Mv polypeptide. For example, such genetic lesions or mutations can be detected by ascertaining the existence of at least one of: 1) a deletion of one or more nucleotides from an SGA-56M or SGA-56Mv gene; 2) an addition of one or more nucleotides to an SGA-56M or SGA-56Mv gene; 3) a substitution of one or more nucleotides of an SGA-56M or SGA-56Mv gene i.e. a point mutation; 4) a chromosomal rearrangement of an SGA-56M or SGA-56Mv gene; 5) an alteration in the level of a messenger RNA transcript of an SGA-56M or SGA-56Mv gene; 6) an aberrant modification of an SGA-56M or SGA-56Mv gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of an SGA-56M or SGA-56Mv gene; 8) a non-wild type level of the protein encoded by an SGA-56M or SGA-56Mv gene; 9) an allelic loss of an SGA-56M or SGA-56Mv gene; and 10) an inappropriate post-translational modification of a protein encoded by an SGA-56M or SGA-56Mv gene. As described herein, there are a large number of assay techniques known in the art that can be used for detecting lesions in a gene.

**[0244]** In certain embodiments, methods for the detection of the lesion involve the use of a probe/primer in a polymerase chain reaction (PCR) (*See, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, *e.g.*, Landegran *et al.* (1988) *Science* 241:1077; and Nakazawa *et al.* (1994) *Proc Natl Acad Sci. USA* 91:360), the latter of which can be particularly useful for detecting point mutations in a gene (*see, e.g.*, Abravaya *et al.* (1995) *Nucleic Acids Res.* 23:675). These methods are useful in the diagnosis and prognosis of cancer in a subject. This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to the selected gene under conditions such that hybridization and amplification of the gene or gene product (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated

that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

**[0245]** Mutations in a selected gene from a sample cell or tissue can also be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, *e.g.*, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

**[0246]** In other embodiments, methods are provided whereby genetic mutations can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays comprising hundreds or thousands of oligonucleotides probes (Cronin et al. 1996, *Human Mutation* 7:244; Kozal et al. 1996, *Nature Medicine* 2:753). For example, genetic mutations can be identified in two-dimensional arrays containing light-generated DNA probes as described in Cronin et al., *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

**[0247]** Sequencing reactions known in the art can be used to directly sequence the selected gene and detect mutations in the SGA-56M or SGA-56Mv gene by comparing the sequence of the sample nucleic acids with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert (Maxim and Gilbert, 1977, *Proc Natl Acad Sci. USA* 74:560) or Sanger (Sanger *et al.* 1977, *Proc Natl Acad Sci. USA* 74:5463). Such methods are useful in the diagnosis and prognosis of a subject with cancer. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve *et al.*, 1995, BioTechniques 19:448), including sequencing by mass spectrometry (see, *e.g.*, PCT Publication No. WO 94/16101; Cohen *et al.* 1996, *Adv. Chromatogr.* 36: 127; and Griffin *et al.*, 1993, *Appl. Biochem. Biotechnol.* 38:147).

**[0248]** Furthermore, the presence of an SGA-56M or SGA-56Mv nucleic acid molecule or polypeptide of the invention can be correlated with the presence or expression level of other cancer-related proteins, such as for example, an androgen receptor, estrogen receptor, adhesion molecules (*e.g.*, E-cadherin), proliferation markers (*e.g.*, MIB-1), tumor-suppressor genes (*e.g.*, TP53, retinoblastoma gene product), vascular endothelial growth factor (Lissoni *et al.*, 2000, *Int J Biol Markers.* 15(4):308), Rad51 (Maacke *et al.*, 2000, *Int J Cancer.* 88(6):907), cyclin D1, BRCA1, BRCA2, or carcinoembryonic antigen.

**[0249]** The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one nucleic acid probe or antibody reagent described herein, which may be

conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a gene encoding a polypeptide of the invention. Furthermore, any cell type or tissue, *e.g.*, preferably cancerous breast cells or tissue, in which the SGA-56M or SGA-56Mv gene is expressed, may be utilized in the prognostic assays described herein.

## 5.5. SCREENING FOR MODULATORS OF SGA-56M AND/OR SGA-56Mv ACTIVITY

**[0250]** The present invention further provides methods for the identification of compounds that may, through their interaction with the SGA-56M and/or SGA-56Mv gene or SGA-56M and/or SGA-56Mv gene product, affect the onset, progression and metastatic spread of breast cancer and/or other cancers.

**[0251]** The following assays are designed to identify: (i) compounds that bind to SGA-56M and/or SGA-56Mv gene products; (ii) compounds that bind to other proteins that interact with an SGA-56M and/or SGA-56Mv gene product; (iii) compounds that interfere with the interaction of the SGA-56M and/or SGA-56Mv gene product with other proteins; and (iv) compounds that modulate the activity of an SGA-56M and/or SGA-56Mv gene (*i.e.*, modulate the level of SGA-56M and/or SGA-56Mv gene expression, including transcription of the SGA-56M and/or SGA-56Mv gene and/or translation of its encoded transcript, and/or modulate SGA-56M and/or SGA-56Mv-encoded polypeptide activity).

**[0252]** Assays may additionally be utilized which identify compounds that bind to SGA-56M and/or SGA-56Mv gene regulatory sequences (*e.g.*, promoter sequences), which may modulate the level of SGA-56M and/or SGA-56Mv gene expression (*see e.g.*, Platt, K.A., 1994, *J. Biol. Chem.* 269:28558).

**[0253]** Such proteins that interact with SGA-56M and/or SGA-56Mv may be involved in the onset, development and metastatic spread of breast cancer or other cancers. Accordingly, methods to modulate the expression level and/or activity of a protein that interacts with SGA-56M and/or SGA-56Mv may also present an effective approach toward modulating the expression and/or activity of SGA-56M and/or SGA-56Mv.

**[0254]** The present invention also provides methods of using isolated SGA-56M and/or SGA-56Mv nucleic acid molecules, or derivatives thereof, as probes that can be used to screen for DNA-binding proteins, including but not limited to proteins that affect DNA conformation or modulate transcriptional activity (*e.g.*, enhancers, transcription factors). In another embodiment, such probes can be used to screen for RNA-binding factors, including but not limited to proteins, steroid hormones, or other small molecules. In yet another embodiment, such probes can be used to detect and identify molecules that bind or affect the pharmacokinetics or activity (*e.g.*, enzymatic activity) of the SGA-56M and/or SGA-56Mv gene or gene product. The protein- or nucleic acid-binding factors or transcriptional modulators identified by a screening assay would provide an appropriate reagent for anti-cancer therapeutics.

**[0255]** In one embodiment, a screening assay of the invention can identify a test compound that is useful for increasing or decreasing the translation of one or both SGA-56M or SGA-56Mv ORFs, for example, by binding to one or more regulatory elements in the 5' untranslated region, the 3' untranslated region, or the coding regions of the mRNA. Compounds that bind to mRNA can, *inter alia*, increase or decrease the rate of mRNA processing, alter its transport through the cell, prevent or enhance binding of the mRNA to ribosomes,

suppressor proteins or enhancer proteins, or alter mRNA stability. Compounds that increase or decrease mRNA translation, for example, can be used to treat or prevent disease. For example, diseases such as cancer, associated with overproduction of proteins, such as SGA-56M and/or SGA-56Mv, can be treated and/or prevented by decreasing translation of the mRNA that codes for the protein, thus inhibiting production of the protein.

**[0256]** Accordingly, in one embodiment, a compound identified by a screening assay of the invention inhibits the production of an SGA-56M and/or SGA-56Mv protein. In a further embodiment, the compound inhibits the translation of an SGA-56M and/or SGA-56Mv mRNA. In yet another embodiment, the compound inhibits transcription of the SGA-56M and/or SGA-56Mv gene.

**[0257]** The invention provides a method for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) which bind to the SGA-56M and/or SGA-56Mv gene product or fragments thereof or have a stimulatory or inhibitory effect on, for example, expression or activity of the SGA-56M and/or SGA-56Mv gene product or fragments thereof.

**[0258]** Compounds identified via assays such as those described herein may be useful, for example, in elaborating the biological function of the SGA-56M and/or SGA-56Mv gene product, and for ameliorating symptoms of breast cancer or other types of cancer. Assays for testing the effectiveness of compounds, identified by, for example, techniques such as those described in Section 5.5.1, are discussed, below, in Section 5.5.3. It is to be noted that the compositions of the invention include pharmaceutical compositions comprising one or more of the compounds identified via such methods. Such pharmaceutical compositions can be formulated, for example, as discussed, below, in Section 5.7.

### 5.5.1. IN VITRO SCREENING ASSAYS FOR COMPOUNDS THAT BIND TO THE SGA-56M AND/OR SGA-56Mv GENE PRODUCT

**[0259]** *In vitro* systems may be designed to identify compounds capable of interacting with, *e.g.*, binding to, an SGA-56M and/or SGA-56Mv gene product of the invention. Compounds identified may be useful, for example, in modulating the activity of wild type and/or mutant SGA-56M and/or SGA-56Mv gene products, may be useful in elaborating the biological function of the SGA-56M and/or SGA-56Mv gene product, may be utilized in screens for identifying compounds that disrupt normal SGA-56M and/or SGA-56Mv gene product interactions, or may in themselves disrupt such interactions. Thus said compounds would be useful in treating, preventing and/or diagnosing cancer. In a particular embodiment said compounds are useful in the treatment, prevention and diagnosis of breast cancer.

**[0260]** The principle of the assays used to identify compounds that interact with the SGA-56M and/or SGA-56Mv gene product involves preparing a reaction mixture of the SGA-56M and/or SGA-56Mv gene product and the test compound under conditions and for a time sufficient to allow the two components to interact with, *e.g.*, bind to, thus forming a complex, which can represent a transient complex, which can be removed and/or detected in the reaction mixture. These assays can be conducted in a variety of ways. For example, one method to conduct such an assay would involve anchoring SGA-56M or SGA-56Mv gene product or the test substance onto a solid phase and detecting SGA-56M or SGA-56Mv gene product/test compound complexes

anchored on the solid phase at the end of the reaction. In one embodiment of such a method, the SGA-56M or SGA-56Mv gene product may be anchored onto a solid surface, and the test compound, which is not anchored, may be labeled, either directly or indirectly.

**[0261]** In practice, microtiter plates may conveniently be utilized as the solid phase. The anchored component may be immobilized by non-covalent or covalent attachments. Non-covalent attachment may be accomplished by simply coating the solid surface with a solution of the protein and drying. Alternatively, an immobilized antibody, preferably a monoclonal antibody, specific or selective for the protein to be immobilized may be used to anchor the protein to the solid surface. The latter method provides for presentation of the protein in a known orientation. The surfaces may be prepared in advance and stored.

**[0262]** In order to conduct the assay, the nonimmobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously nonimmobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously nonimmobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; *e.g.*, using a labeled antibody specific for the previously nonimmobilized component (the antibody, in turn, may be directly labeled or indirectly labeled with a labeled anti-Ig antibody).

**[0263]** Alternatively, a reaction can be conducted in a liquid phase, the reaction products separated from unreacted components, and complexes detected; *e.g.*, using an immobilized antibody specific for an SGA-56M and/or SGA-56Mv gene product or the test compound to anchor any complexes formed in solution, and a labeled antibody specific for the other component of the possible complex to detect anchored complexes.

### 5.5.2 ASSAYS FOR PROTEINS THAT INTERACT WITH AN SGA-56M OR SGA-56Mv GENE PRODUCT

**[0264]** Any method suitable for detecting protein-protein interactions may be employed for identifying SGA-56M and/or SGA-56Mv protein-protein interactions. Proteins that interact with SGA-56M and/or SGA-56Mv will be potential therapeutics for the treatment of cancer. Thus the assays described below are useful in identifying proteins that can be used in methods to treat cancer. Proteins that interact with SGA-56M and/or SGA-56Mv can also be used in the diagnosis of cancer. Thus, the assays described below are also useful in methods to diagnose cancer.

**[0265]** Among the traditional methods that may be employed are co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns (*e.g.*, size exclusion chromatography). Utilizing procedures such as these allows for the isolation of intracellular proteins that interact with SGA-56M and/or SGA-56Mv gene products. Once isolated, such an intracellular protein can be identified and can, in turn, be used, in conjunction with standard techniques, to identify additional proteins with which it interacts. For example, at least a portion of the amino acid sequence of an intracellular protein or a protein having an

intracellular domain which interacts with the SGA-56M and/or SGA-56Mv gene product can be ascertained using techniques well known to those of skill in the art, such as via the Edman degradation technique (*see, e.g.*, Creighton, 1983, *Proteins: Structures and Molecular Principles*, W.H. Freeman & Co., N.Y., pp.34-49). The amino acid sequence obtained may be used as a guide for the generation of oligonucleotide mixtures that can be used to screen for gene sequences encoding such proteins. Screening may be accomplished, for example, by standard hybridization or PCR techniques. Techniques for the generation of oligonucleotide mixtures and screening are well known. (*See, e.g.*, Ausubel, *supra.*, and *PCR Protocols: A Guide to Methods and Applications*, 1990, Innis, M. *et al.*, eds. Academic Press, Inc., New York).

**[0266]** Additionally, methods may be employed which result in the simultaneous identification of genes which encode a protein interacting with the SGA-56M and/or SGA-56Mv protein. These methods include, for example, probing expression libraries with labeled SGA-56M or SGA-56Mv protein in a manner similar to the well known technique of antibody probing of λgt11 libraries.

**[0267]** One method that detects protein interactions *in vivo*, the two-hybrid system, may also be used. Many versions of this system have been described (see, e.g., Chien *et al.*, 1991, *supra*) and some are commercially available from Clontech (Palo Alto, CA).

### 5.5.3. ASSAYS FOR COMPOUNDS THAT INTERFERE WITH SGA-56M AND/OR SGA-56Mv INTERACTION

**[0268]** The SGA-56M and/or SGA-56Mv gene product may, *in vivo*, interact with one or more macromolecules, such as proteins or nucleic acids. Such macromolecules are referred to herein as "interacting partners" or "specific binding partners". Compounds that disrupt an association of SGA-56M and/or SGA-56Mv with interacting partner(s) may be useful in regulating the activity of the SGA-56M and/or SGA-56Mv gene product, including mutant SGA-56M and/or SGA-56Mv gene products. Such compounds may include, but are not limited to molecules such as peptides, and the like, as described, for example, in Section 5.5.1., which would be capable of interacting with SGA-56M and/or SGA-56Mv polypeptides. Thus the assays described below are useful for identifying proteins and/or nucleic acids that can be used in methods to treat cancer. Proteins and nucleic acids that interact with SGA-56M and/or SGA-56Mv can also be used in the diagnosis of cancer, *e.g.*, breast cancer. Thus the assays described below are also useful in methods to diagnose cancer, *e.g.*, breast cancer.

**[0269]** The basic principle of the assay systems used to identify compounds that interfere with the interaction between the SGA-56M and/or SGA-56Mv gene product and an interacting partner or partners involves preparing a reaction mixture containing the SGA-56M and/or SGA-56Mv gene product, and the interacting partner under conditions and for a time sufficient to allow the two to interact and bind, thus forming a complex. In order to test a compound for inhibitory activity, the reaction mixture is prepared in the presence and absence of the test compound. The test compound may initially be included in the reaction mixture, or may be added at a time subsequent to the addition of SGA-56M and/or SGA-56Mv gene product and an interacting partner. Control reaction mixtures are incubated without the test compound or with a control compound. The formation of any complexes between an SGA-56M and/or SGA-56Mv gene protein and an

interacting partner is then detected. The formation of a complex in the control reaction, but not in a reaction mixture comprising a test compound, indicates that the compound interferes with the interaction of the SGA-56M and/or SGA-56Mv gene product and the interacting partner. Additionally, complex formation within reaction mixtures containing the test compound and normal SGA-56M and/or SGA-56Mv gene protein may also be compared to complex formation within reaction mixtures containing the test compound and a mutant form of either an SGA-56M and/or SGA-56Mv geneproduct. This comparison may be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal SGA-56M and/or SGA-56Mv gene proteins.

**[0270]** The assay for compounds that interfere with the interaction of the SGA-56M and/or SGA-56Mv gene products and interacting partners can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either an SGA-56M or SGA-56Mv gene product or the binding partner onto a solid phase and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the SGA-56M or SGA-56Mv gene products and the interacting partners, *e.g.*, by competition, can be identified by conducting the reaction in the presence of the test substance; *i.e.*, by adding the test substance to the reaction mixture prior to or simultaneously with the SGA-56M or SGA-56Mv gene protein and intracellular interacting partner. Alternatively, test compounds that disrupt preformed complexes, *e.g.*, compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are described briefly below.

**[0271]** In a heterogeneous assay system, either the SGA-56M or SGA-56Mv gene product or the interacting partner, is anchored onto a solid surface, while the non-anchored species is labeled, either directly or indirectly. In practice, microtiter plates are conveniently utilized. The anchored species may be immobilized by non-covalent or covalent attachments. Non-covalent attachment may be accomplished simply by coating the solid surface with a solution of the SGA-56M or SGA-56Mv gene product or interacting partner and drying. Alternatively, an immobilized antibody, for example, specific for the species to be anchored may be used to anchor the species to the solid surface. The surfaces may be prepared in advance and stored.

**[0272]** In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) and any complexes formed remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; *e.g.*, using a labeled antibody specific for the initially non-immobilized species (the antibody, in turn, may be directly labeled or indirectly labeled with, for example, a labeled anti-Ig antibody).

Depending upon the order of addition of reaction components, test compounds which inhibit complex formation or which disrupt preformed complexes can be detected.

**[0273]** Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components, and complexes detected. The reaction, for example, may be executed using an immobilized antibody specific for one of the interacting components to anchor any complexes formed in solution, and the labeled antibody specific for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds which inhibit complex or which disrupt preformed complexes can be identified.

**[0274]** In an alternate embodiment of the invention, a homogeneous assay can be used. In this approach, a preformed complex of the SGA-56M or SGA-56Mv gene protein and the interacting partner is prepared in which either the SGA-56M or SGA-56Mv gene product or its interacting partner is labeled, but the signal generated by the label is quenched due to complex formation (*see, e.g.*, U.S. Patent No. 4,109,496 by Rubenstein). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt SGA-56M and/or SGA-56Mv gene product association with an interacting partner can be identified.

**[0275]** In a particular embodiment, the SGA-56M or SGA-56Mv gene product can be prepared for immobilization using recombinant DNA techniques described in Section 5.1, above. For example, the SGA-56M or SGA-56Mv coding region can be fused to a glutathione-S-transferase (GST) gene using a fusion vector, such as pGEX-5X-1, in such a manner that its interacting activity is maintained in the resulting fusion protein. The interacting partner can be purified and used to raise a monoclonal antibody, using methods routinely practiced in the art and described above, in Section 5.2. This antibody can be labeled with the radioactive isotope $^{125}I$, for example, by methods routinely practiced in the art. In a heterogeneous assay, *e.g.*, the GST-SGA-56M or GST-SGA-56Mv fusion protein can be anchored to glutathione-agarose beads. The interacting partner can then be added in the presence or absence of the test compound in a manner that allows interaction, *e.g.*, binding, to occur. At the end of the reaction period, unbound material can be washed away, and the labeled monoclonal antibody can be added to the system and allowed to bind to the complexed components. The interaction between the SGA-56M or SGA-56Mv gene protein and the interacting partner can be detected by measuring the amount of radioactivity that remains associated with the glutathione-agarose beads. Alternative means may be applied for such approaches, including the generation of detectable and distinguishable fusion proteins comprising each of the interacting proteins, e.g., SGA-56M-GST and a His tagged version of an SGA-56M interacting protein. A successful inhibition of the interaction by the test compound will result in a decrease in measured radioactivity.

**[0276]** Alternatively, the GST-SGA-56M or GST-SGA-56Mv fusion protein and the intracellular interacting partner can be mixed together in liquid in the absence of the solid glutathione-agarose beads. The test compound can be added either during or after the species are allowed to interact. This mixture can then be added to the glutathione-agarose beads and unbound material is washed away. The extent of inhibition of

SGA-56M or SGA-56Mv gene product/binding partner interaction can be detected by addition of a labeled antibody, for example, and measuring the radioactivity associated with the beads.

**[0277]** It will be appreciated that the above assays may be preformed with a mixture of SGA-56M and SGA-56Mv gene product. The ratio at which the different gene products are mixed may be varied according to the application.

### 5.5.4. CELL-BASED ASSAYS FOR SGA-56M OR SGA-56Mv ACTIVITY

**[0278]** Cell-based methods are presented herein which identify compounds capable of treating breast cancer and other cancers by modulating SGA-56M and/or SGA-56Mv activity or expression levels. Specifically, such assays identify compounds that affect SGA-56M and/or SGA-56Mv dependent processes, such as, but not limited to changes in cell morphology, cell division, differentiation, adhesion, motility, phosphorylation, or dephosphorylation of cellular proteins. Such assays can also be used to identify compounds that affect SGA-56M and/or SGA-56Mv expression levels or gene activity directly. Compounds identified via such methods can, for example, be utilized in methods for treating breast cancer and other cancers and metastasis thereof.

**[0279]** In one embodiment, an assay is a cell-based assay in which a cell that expresses a membrane-bound form of the SGA-56M or SGA-56Mv gene product, or a biologically active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to the polypeptide determined. In another embodiment the SGA-56M or SGA-56Mv gene product is cytosolic. The cell, for example, can be a yeast cell or a cell of mammalian origin. Determining the ability of the test compound to bind to the polypeptide can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the polypeptide or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with $^{125}I$, $^{35}S$, $^{14}C$, or $^{3}H$, either directly or indirectly, and the radioisotope detected by direct counting of radio-emission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In a preferred embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or a biologically active portion thereof as compared to a reference compound.

**[0280]** In another embodiment, the cell-based assays are based on expression of the SGA-56M or SGA-56Mv gene product in a mammalian cell and measuring the SGA-56M and/or SGA-56Mv -dependent process. Any mammalian cells that can express the SGA-56M and/or SGA-56Mv gene and allow the functioning of

the SGA-56M and/or SGA-56Mv gene product can be used, in particular, cancer cells derived from the breast, such as MCF-7, BT483, Hs578T, HTB26, BT20 and T47D. Normal mammary gland cell lines such as, for example, CRL7030 and Hs578Bst, may also be used provided that an SGA-56M and/or SGA-56Mv gene product is produced. Other mammalian cell lines that can be used include, but are not limited to, CHO, HeLa, NIH3T3, and Vero cells. Recombinant expression of the SGA-56M and/or SGA-56Mv gene in these cells can be achieved by methods described in Section 5.2. In these assays, cells producing functional SGA-56M and/or SGA-56Mv gene products are exposed to a test compound for an interval sufficient for the compound to modulate the activity of the SGA-56M and/or SGA-56Mv gene product. The activity of SGA-56M and/or SGA-56Mv gene product can be measured directly or indirectly through the detection or measurement of SGA-56M and/or SGA-56Mv-dependent cellular processes. As a control, a cell not producing the SGA-56M and/or SGA-56Mv gene product may be used for comparisons. Depending on the cellular process, any techniques known in the art may be applied to detect or measure it.

**[0281]** In another embodiment a cell or cell line that is capable of expressing SGA-56M and/or SGA-56Mv is contacted with a test compound that is believed to modulate expression of the SGA-56M and/or SGA-56Mv gene. Expression levels of the SGA-56M and/or SGA-56Mv gene can be monitored in the presence or absence of the test compound. Alternatively, expression levels can be monitored in the presence of a test compound as compared to expression levels of the SGA-56M and/or SGA-56Mv gene in the presence of a control compound or a placebo. Any method known in the art can be used to monitor SGA-56M and/or SGA-56Mv gene expression. As an example, but not as a limitation, such methods can include analysis by Western blot, Northern blot, and real-time quantitative RT-PCR.

**[0282]** In yet another embodiment, cells which express the SGA-56M and/or SGA-56Mv gene product, *e.g.*, MCF-7 cells are made permeable, *e.g.*, by treatment with a mild detergent and exposed to a test compound. Binding of the test compound can be detected directly (*e.g.*, radioactively labeling the test compound) or indirectly (e.g., antibody detection) or by any means known in the art.

**[0283]** Any compound can be used in a cell-based assay to test if it affects SGA-56M and/or SGA-56Mv activity or expression levels. The compound can be a protein, a peptide, a nucleic acid, an antibody or fragment thereof, a small molecule, an organic molecule or an inorganic molecule. (*e.g.*, steroid, pharmaceutical drug). A small molecule is considered a non-peptide compound with a molecular weight of less than 500 daltons.

## 5.6. METHODS FOR TREATMENT OF CANCER

**[0284]** Described below are methods and compositions for treating cancer, *e.g.*, breast or lung cancer, using the SGA-56M and/or SGA-56Mv gene or gene product as a therapeutic target. The outcome of a treatment is to at least produce in a treated subject a healthful benefit, which in the case of cancer, including breast cancer, includes but is not limited to remission of the cancer, palliation of the symptoms of the cancer, and/or control of metastatic spread of the cancer.

[0285] All such methods comprise methods for modulating SGA-56M and/or SGA-56Mv gene activity and/or expression which in turn modulate the phenotype of the treated cell and tumorigenic potential.

[0286] As discussed, above, successful treatment of breast cancer or other cancers can be effected through techniques that serve to decrease SGA-56M and/or SGA-56Mv activity. Activity can be decreased, for example, bydirectly decreasing SGA-56M and/or SGA-56Mv gene product activity and/or by decreasing the level of SGA-56M and/or SGA-56Mv gene expression. Thus the invention provides methods of treating a subject with cancer by administering to said subject a therapeutically effective amount of a compound that antagonizes an SGA-56M and/or SGA-56Mv gene product.

[0287] For example, compounds such as those identified through assays described, above, in Section 5.5, above, which decrease SGA-56M and/or SGA-56Mv activity can be used in accordance with the invention to treat breast cancer or other cancers. As discussed in Section 5.5, above, such molecules can include, but are not limited to proteins, nucleic acids, peptides, including soluble peptides, and small organic or inorganic molecules, and can be referred to as SGA-56M and/or SGA-56Mv antagonists. Techniques for the determination of effective doses and administration of such compounds are described, below, in Section 5.7.

[0288] Further, antisense and ribozyme molecules which inhibit SGA-56M and/or SGA-56Mv gene expression can also be used in accordance with the invention to reduce the level of SGA-56M and/or SGA-56Mv gene expression, thus effectively reducing the level of SGA-56M and/or SGA-56Mv gene product present, thereby decreasing the level of SGA-56M and/or SGA-56Mv activity. The invention therefore relates to a pharmaceutical composition comprising an antisense or ribozyme molecule with specificity for an SGA-56M and/or SGA-56Mv gene product. Still further, triple helix molecules can be utilized in reducing the level of SGA-56M and/or SGA-56Mv gene activity. Such molecules can be designed to reduce or inhibit either wild type, or if appropriate, mutant target gene activity. Small organic or inorganic molecules can also be used to inhibit SGA-56M and/or SGA-56Mv gene expression and/or inhibit production and/or activity of an SGA-56M and/or SGA-56Mv gene product. Techniques for the production and use of such molecules are well known to those of skill in the art.

### 5.6.1. ANTISENSE MOLECULES

[0289] Anti-sense nucleic acid molecules which are complementary to nucleic acid sequences contained within the SGA-56M or SGA-56Mv gene as shown in FIG. 1 (SEQ ID NO: 1) and FIG. 2 (SEQ ID NO: 3), including but not limited to anti-sense nucleic acid molecules complementary to (SEQ ID NO: 1) and (SEQ ID NO: 3), can be used to treat any cancer, in which the expression level of the SGA-56M or SGA-56Mv gene is elevated in cancerous cells or tissue as compared to normal cells or tissue or a predetermined non-cancerous standard. Thus in one embodiment of the invention a method of treating breast cancer is provided whereby a patient suffering from breast cancer is treated with a therapeutically effective amount of an SGA-56M or SGA-56Mv anti-sense nucleic acid molecule.

[0290] Antisense approaches involve the design of oligonucleotides (either DNA or RNA) that are complementary to SGA-56M or SGA-56Mv gene mRNA. The antisense oligonucleotides will bind to the

complementary SGA-56M or SGA-56Mv gene mRNA transcripts and prevent translation. Absolute complementarity, although preferred, is not required. A sequence "complementary" to a portion of an RNA, as referred to herein, means a sequence having sufficient complementarity to be able to hybridize with the non-poly A portion of the RNA, forming a stable duplex; in the case of double-stranded antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The ability to hybridize will depend on both the degree of complementarity and the length of the antisense nucleic acid. Generally, the longer the hybridizing nucleic acid, the more base mismatches with regard to a nucleic acid target it may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

**[0291]** Oligonucleotides that are complementary to the 5' end of the message, *e.g.*, the 5' untranslated sequence up to and including the AUG initiation codon, should work most efficiently at inhibiting translation. However, sequences complementary to the 3' untranslated sequences of mRNAs have also been shown to be effective at inhibiting translation of mRNAs as well. (See generally, Wagner, 1994, *Nature* 372:333). Thus, oligonucleotides complementary to the 5'-non-translated region, the 3'-non-translated region, or the non-translated, non-coding region between the SGA-56M or SGA-56Mv open reading frame of the SGA-56M or SGA-56Mv gene (referred to herein after as the "intervening region", as shown, for example, in FIG. 1, could be used in an antisense approach to inhibit translation of endogenous SGA-56M or SGA-56Mv gene mRNA.

**[0292]** Oligonucleotides complementary to the 5' untranslated region of the mRNA should ideally include the complement of the AUG start codon. Antisense oligonucleotides complementary to mRNA coding regions are less efficient inhibitors of translation but could be used in accordance with the invention. Whether designed to hybridize to the 5'-, 3'-, intervening, or coding region of SGA-56M and/or SGA-56Mv gene mRNA, antisense nucleic acids should be at least six nucleotides in length, and are preferably oligonucleotides ranging from 6 to about 50 nucleotides in length. In specific aspects the oligonucleotide is at least 10 nucleotides, at least 17 nucleotides, at least 25 nucleotides or at least 50 nucleotides.

**[0293]** Regardless of the choice of target sequence, it is preferred that *in vitro* studies are first performed to quantitate the ability of the antisense oligonucleotide to inhibit gene expression. It is preferred that these studies utilize controls that distinguish between antisense gene inhibition and nonspecific biological effects of oligonucleotides. It is also preferred that these studies compare levels of the target RNA or protein with that of an internal control RNA or protein. Additionally, it is envisioned that results obtained using the antisense oligonucleotide are compared to those obtained using a control oligonucleotide. It is preferred that the control oligonucleotide is of approximately the same length as the test oligonucleotide and that the nucleotide sequence of the oligonucleotide differs from the antisense sequence no more than is necessary to prevent specific hybridization to the target sequence.

**[0294]** The oligonucleotides can be DNA or RNA or chimeric mixtures or derivatives or modified versions thereof, single-stranded or double-stranded. The oligonucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone, for example, to improve stability of the molecule, hybridization, etc. The

oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. USA* 86:6553; Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci. USA* 84:648; PCT Publication No. WO88/09810, published December 15, 1988) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO89/10134, published April 25, 1988), hybridization-triggered cleavage agents. (*see, e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958) or intercalating agents. (*see, e.g.*, Zon, 1988, *Pharm. Res.* 5:539). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

**[0295]** The antisense oligonucleotide may comprise at least one modified base moiety which is selected from the group including but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

**[0296]** The antisense oligonucleotide may also comprise at least one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose.

**[0297]** In yet another embodiment, the antisense oligonucleotide comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

**[0298]** In yet another embodiment, the antisense oligonucleotide is an α-anomeric oligonucleotide. An α-anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gautier *et al.*, 1987, *Nucl. Acids Res.* 15:6625). The oligonucleotide is a 2'-0-methylribonucleotide (Inoue *et al.*, 1987, *Nucl. Acids Res.* 15:6131), or a chimeric RNA-DNA analogue (Inoue *et al.*, 1987, *FEBS Lett.* 215:327).

**[0299]** The SGA-56M antisense nucleic acid sequence can comprise the complement of any contiguous segment within the sequence of the SGA-56M gene (SEQ ID NO: 1).

**[0300]** In one embodiment of the present invention, the SGA-56M antisense nucleic acid sequence is about 50 bp in length. In certain specific embodiments, the SGA-56M antisense nucleic acid sequence comprises the sequence complementary to any contiguous block of 50, 100, 200, or 400 nucleotides of SEQ ID NO: 1.

**[0301]** The SGA-56Mv antisense nucleic acid sequence can comprise the complement of any contiguous segment within the sequence of the SGA-56Mv gene (SEQ ID NO: 3).

**[0302]** In one embodiment of the present invention, the SGA-56Mv antisense nucleic acid sequence is about 50 bp in length. In certain specific embodiments, the SGA-56Mv antisense nucleic acid sequence comprises the sequence complementary to any contiguous block of 50, 100, 200, or 400 nucleotides of SEQ ID NO: 3.

**[0303]** Oligonucleotides of the invention may be synthesized by standard methods known in the art, *e.g.*, by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, *etc.*). As examples, phosphorothioate oligonucleotides may be synthesized by the method of Stein *et al.* (1988, *Nucl. Acids Res.* 16:3209), methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin *et al.*, 1988, *Proc. Natl. Acad. Sci. U.S.A.* 85:7448), *etc.*

**[0304]** While antisense nucleotides complementary to the SGA-56M coding region could be used, those complementary to the transcribed untranslated region are most preferred.

**[0305]** The antisense molecules should be delivered to cells that express the SGA-56M and/or SGA-56Mv gene *in vivo*. A number of methods have been developed for delivering antisense DNA or RNA to cells; *e.g.*, antisense molecules can be injected directly into the tissue site, or modified antisense molecules, designed to target the desired cells (*e.g.*, antisense linked to peptides or antibodies that specifically bind receptors or antigens expressed on the target cell surface) can be administered systemically.

**[0306]** However, it is often difficult to achieve intracellular concentrations of the antisense sufficient to suppress translation of endogenous mRNAs. Therefore a preferred approach utilizes a recombinant DNA construct in which the antisense oligonucleotide is placed under the control of a strong pol III or pol II promoter. The use of such a construct to transfect target cells in a patient results in the transcription of sufficient amounts of single stranded RNAs that will form complementary base pairs with the endogenous SGA-56M and/or SGA-56Mv gene transcripts and thereby prevent translation of the SGA-56M and/or SGA-56Mv gene mRNA. For example, a vector can be introduced *in vivo* such that it can be taken up by a cell and direct the transcription of an antisense RNA. Such a vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art. Vectors can be a plasmid, viral vector, or other construct known in the art, used for replication and expression in mammalian cells. Expression of the sequence encoding the antisense RNA can be by any promoter known in the art to function in mammalian, preferably human cells. Such promoters can be inducible or constitutive. Such promoters include but are not limited to: the SV40 early promoter region (Bernoist and Chambon, 1981, *Nature* 290:304), the promoter contained in the 3′ long terminal repeat of Rous sarcoma virus (Yamamoto *et al.*, 1980, *Cell* 22:787), the herpes thymidine kinase promoter (Wagner *et al.*, 1981, *Proc. Natl. Acad. Sci. USA* 78:1441), the regulatory sequences of the metallothionein gene (Brinster *et al.*, 1982, *Nature* 296:39), etc. Any type of plasmid, cosmid, YAC or viral vector can be used to prepare the recombinant DNA construct that can be introduced directly into the tissue site. Alternatively, viral vectors can be used which selectively infect the desired tissue.

**[0307]** The effective dose of SGA-56M and/or SGA-56Mv antisense oligonucleotide to be administered during a treatment cycle ranges from about 0.01 to 0.1, 0.1 to 1, or 1 to 10 mg/kg/day. The dose of SGA-56M

and/or SGA-56Mv antisense oligonucleotide to be administered may depend on the mode of administration. For example, intravenous administration of an SGA-56M and/or SGA-56Mv antisense oligonucleotide would likely result in a significantly higher systemic dose than a full body dose resulting from a local implant comprising a pharmaceutical composition comprising SGA-56M and/or SGA-56Mv antisense oligonucleotide. In one embodiment, an SGA-56M and/or SGA-56Mv antisense oligonucleotide is administered subcutaneously at a dose of 0.01 to 10 mg/kg/day. In another embodiment, an SGA-56M and/or SGA-56Mv antisense oligonucleotide is administered intravenously at a dose of 0.01 to 10 mg/kg/day. In yet another embodiment, an SGA-56M and/or SGA-56Mv antisense oligonucleotide is administered locally at a dose of 0.01 to 10 mg/kg/day. It will be evident to one skilled in the art that local administrations can result in lower total body doses. For example, local administration methods such as intratumor administration, intraocular injection, or implantation, can produce locally high concentrations of SGA-56M and/or SGA-56Mv antisense oligonucleotide, but represent a relatively low dose with respect to total body weight. Thus, in such cases, local administration of an SGA-56M and/or SGA-56Mv antisense oligonucleotide is contemplated to result in a total body dose of about 0.01 to 5 mg/kg/day.

**[0308]** In another embodiment, a particularly high dose of SGA-56M and/or SGA-56Mv antisense oligonucleotide, which ranges from about 10 to 50 mg/kg/day, is administered during a treatment cycle.

**[0309]** Moreover, the effective dose of a particular SGA-56M and/or SGA-56Mv antisense oligonucleotide may depend on additional factors, including the type of disease, the disease state or stage of disease, oligonucleotide toxicity, oligonucleotide stability, the oligonucleotide's rate of uptake by cancer cells, as well as the weight, age, and health of the individual to whom the antisense oligonucleotide is to be administered. Because of the many factors present *in vivo* that may interfere with the action or biological activity of an SGA-56M and/or SGA-56Mv antisense oligonucleotide, one of ordinary skill in the art can appreciate that an effective amount of an SGA-56M and/or SGA-56Mv antisense oligonucleotide may vary for each individual.

**[0310]** In another embodiment, an SGA-56M and/or SGA-56Mv antisense oligonucleotide is at a dose which results in circulating plasma concentrations of an SGA-56M and/or SGA-56Mv antisense oligonucleotide which is at least 50 nM (nanomolar). As will be apparent to the skilled artisan, lower or higher plasma concentrations of an SGA-56M and/or SGA-56Mv antisense oligonucleotide may be preferred depending on the mode of administration. For example, plasma concentrations of an SGA-56M and/or SGA-56Mv antisense oligonucleotide of at least 50 nM can be appropriate in connection with intravenous, subcutaneous, intramuscular, controlled release, and oral administration methods, to name a few. In another example, relatively low circulating plasma levels of an SGA-56M and/or SGA-56Mv antisense oligonucleotide can be desirable, however, when using local administration methods such as, for example, intratumor administration, intraocular administration, or implantation, which nevertheless can produce locally high, clinically effective concentrations of SGA-56M and/or SGA-56Mv antisense oligonucleotide.

**[0311]** The high dose may be achieved by several administrations per cycle. Alternatively, the high dose may be administered in a single bolus administration. A single administration of a high dose may result in

circulating plasma levels of SGA-56M and/or SGA-56Mv antisense oligonucleotide that are transiently much higher than 50 nM.

**[0312]** Additionally, the dose of an SGA-56M and/or SGA-56Mv antisense oligonucleotide may vary according to the particular SGA-56M and/or SGA-56Mv antisense oligonucleotide used. The dose employed is likely to reflect a balancing of considerations, among which are stability, localization, cellular uptake, and toxicity of the particular SGA-56M and/or SGA-56Mv antisense oligonucleotide. For example, a particular chemically modified SGA-56M and/or SGA-56Mv antisense oligonucleotide may exhibit greater resistance to degradation, or may exhibit higher affinity for the target nucleic acid, or may exhibit increased uptake by the cell or cell nucleus, any of which properties may permit the use of lower doses. In yet another example, a particular chemically modified SGA-56M and/or SGA-56Mv antisense oligonucleotide may exhibit lower toxicity than other antisense oligonucleotides, and therefore can be used at high doses. Thus, for a given SGA-56M and/or SGA-56Mv antisense oligonucleotide, an appropriate dose to administer can be relatively high or relatively low. The invention contemplates the continued assessment of optimal treatment schedules for particular species of SGA-56M and/or SGA-56Mv antisense oligonucleotides. The daily dose can be administered in one or more treatments.

**[0313]** A "low dose" or "reduced dose" refers to a dose that is below the normally administered range, *i.e.*, below the standard dose as suggested by the Physicians' Desk Reference, 54th Edition (2000) or a similar reference. Such a dose can be sufficient to inhibit cell proliferation, demonstrate ameliorative effects in a human, or demonstrate efficacy with fewer side effects as compared to standard cancer treatments. Normal dose ranges used for particular therapeutic agents and standard cancer treatments employed for specific diseases can be found in the Physicians' Desk Reference, 54th Edition (2000) or in Cancer: Principles & Practice of Oncology, DeVita, Jr., Hellman, and Rosenberg (eds.) 2nd edition, Philadelphia, PA: J.B. Lippincott Co., 1985.

**[0314]** Reduced doses of an SGA-56M or SGA-56Mv nucleic acid molecule, SGA-56M or SGA-56Mv polypeptide, SGA-56M and/or SGA-56Mv antagonist, and/or combination therapeutic can demonstrate reduced toxicity, such that fewer side effects and toxicities are observed in connection with administering an SGA-56M and/or SGA-56Mv antagonist and one or more cancer therapeutics for shorter duration and/or at lower dosages when compared to other treatment protocols and dosage formulations, including the standard treatment protocols and dosage formulations as described in the Physicians' Desk Reference, 54th Edition (2000) or in Cancer: Principles & Practice of Oncology, DeVita, Jr., Hellman, and Rosenberg (eds.) 2nd edition, Philadelphia, PA: J.B. Lippincott Co., 1985.

**[0315]** A "treatment cycle" or "cycle" refers to a period during which at least one therapeutic or sequence of therapeutics is administered. In some instances, one treatment cycle may be desired, such as, for example, in the case where a significant therapeutic effect is obtained after one treatment cycle. The present invention contemplates at least one treatment cycle, generally preferably more than one treatment cycle.

**[0316]** Other factors to be considered in determining an effective dose of an SGA-56M and/or SGA-56Mv antisense oligonucleotide include whether the oligonucleotide will be administered in combination with other

therapeutics. In such cases, the relative toxicity of the other therapeutics may indicate the use of an SGA-56M and/or SGA-56Mv antisense oligonucleotide at low doses. Alternatively, treatment with a high dose of SGA-56M and/or SGA-56Mv antisense oligonucleotide can result in combination therapies with reduced doses of therapeutics. In a specific embodiment, treatment with a particularly high dose of SGA-56M and/or SGA-56Mv antisense oligonucleotide can result in combination therapies with greatly reduced doses of cancer therapeutics. For example, treatment of a patient with 10, 20, 30, 40, or 50 mg/kg/day of an SGA-56M and/or SGA-56Mv antisense oligonucleotide can further increase the sensitivity of a subject to cancer therapeutics. In such cases, the particularly high dose of SGA-56M and/or SGA-56Mv antisense oligonucleotide is combined with, for example, a greatly shortened radiation therapy schedule. In another example, the particularly high dose of an SGA-56M and/or SGA-56Mv antisense oligonucleotide produces significant enhancement of the potency of cancer therapeutic agents.

**[0317]** Additionally, the particularly high doses of SGA-56M and/or SGA-56Mv antisense oligonucleotide may further shorten the period of administration of a therapeutically effective amount of SGA-56M and/or SGA-56Mv antisense oligonucleotide and/or additional therapeutic, such that the length of a treatment cycle is much shorter than that of the standard treatment.

**[0318]** The invention contemplates other treatment regimens depending on the particular SGA-56M and/or SGA-56Mv antisense oligonucleotide to be used, or depending on the particular mode of administration, or depending on whether an SGA-56M and/or SGA-56Mv antisense oligonucleotide is administered as part of a combination therapy, *e.g.*, in combination with a cancer therapeutic agent. The daily dose can be administered in one or more treatments.

### 5.6.2. RIBOZYME MOLECULES

**[0319]** Ribozyme molecules that are complementary to RNA sequences encoded for by the SGA-56M or SGA-56Mv gene as shown in FIG. 1 and FIG. 2 can be used to treat any cancer, including breast cancer. Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA (for a review see, for example Rossi, J., 1994, *Current Biology* 4:469). The mechanism of ribozyme action involves sequence specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Ribozyme molecules include one or more sequences complementary to the target gene mRNA, and the well known catalytic sequence responsible for mRNA cleavage (*See* U.S. Pat. No. 5,093,246). As such, within the scope of the invention are engineered hammerhead motif ribozyme molecules that specifically and efficiently catalyze endonucleolytic cleavage of RNA sequences encoding target gene proteins. Ribozyme molecules designed to catalytically cleave SGA-56M or SGA-56Mv mRNA transcripts can also be used to prevent translation of SGA-56M or SGA-56Mv mRNA. (*See, e.g.*, PCT International Publication WO90/11364, published October 4, 1990; Sarver *et al.*, 1990, *Science* 247:1222). While ribozymes that cleave mRNA at site-specific recognition sequences can be used to destroy SGA-56M or SGA-56Mv mRNAs, the use of hammerhead ribozymes is preferred. Hammerhead ribozymes cleave mRNAs at locations dictated by flanking regions that form complementary base pairs with the target mRNA. The sole

requirement is that the target mRNA have the following sequence of two bases: 5'-UG-3'. The construction and production of hammerhead ribozymes is well known in the art and is described more fully in Haseloff and Gerlach, 1988, *Nature* 334:585. Preferably the ribozyme is engineered so that the cleavage recognition site is located near the 5' end of the SGA-56M or SGA-56Mv mRNA; *i.e.*, to increase efficiency and minimize the intracellular accumulation of non-functional mRNA transcripts.

**[0320]** The ribozymes of the present invention also include RNA endoribonucleases (hereinafter "Cech-type ribozymes") such as the one which occurs naturally in *Tetrahymena Thermophila* (known as the IVS, or L-19 IVS RNA) and which has been extensively described by Cech and collaborators (Zaug *et al.*, 1984, *Science* 224:574; Zaug and Cech, 1986, *Science* 231:470; Zaug *et al.*, 1986, *Nature* 324:429; published International patent application No. WO 88/04300 by University Patents Inc.; Been and Cech, 1986, *Cell* 47:207). The Cech-type ribozymes have an eight base pair active site that hybridizes to a target RNA sequence whereafter cleavage of the target RNA takes place. The invention encompasses those Cech-type ribozymes that target eight base-pair active site sequences that are present in an SGA-56M or SGA-56Mv gene.

**[0321]** As in the antisense approach, the ribozymes can be composed of modified oligonucleotides (*e.g.*, for improved stability, targeting, etc.) and should be delivered to cells that express the SGA-56M or SGA-56Mv gene *in vivo*. A preferred method of delivery involves using a DNA construct "encoding" the ribozyme under the control of a strong constitutive pol III or pol II promoter, so that transfected cells will produce sufficient quantities of the ribozyme to destroy endogenous SGA-56M or SGA-56Mv transcripts and inhibit translation. Ribozymes, unlike antisense molecules, are catalytic and require a lower intracellular concentration to ensure effectiveness.

**[0322]** Anti-sense RNA and DNA, ribozyme, and triple helix molecules of the invention can be prepared by any method known in the art for the synthesis of DNA and RNA molecules. These include techniques for chemically synthesizing oligodeoxyribonucleotides and oligoribonucleotides well known in the art such as for example solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules can be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences can be incorporated into a wide variety of vectors that incorporate suitable RNA polymerase promoters such as T7 or SP6 polymerase promoters. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

**[0323]** Various well-known modifications to the DNA molecules can be introduced as a means of increasing intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences of ribo- or deoxy- nucleotides to the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the oligodeoxyribonucleotide backbone.

### 5.6.3. THERAPEUTIC ANTIBODIES

**[0324]** Antibodies exhibiting capability to downregulate SGA-56M and/or SGA-56Mv gene product activity can be utilized to treat breast cancer and other cancers wherein SGA-56M and/or SGA-56Mv expression levels are elevated. Antibodies immunologically specific for wild type or mutant SGA-56M and/or SGA-56Mv proteins, or peptides corresponding to portions of the proteins can be generated using standard techniques described in Section 5.3. Such antibodies include, but are not limited to polyclonal, monoclonal, Fab fragments, single chain antibodies, chimeric antibodies, and the like.

**[0325]** Antibodies that recognize any epitope on the SGA-56Mand/or SGA-56Mv protein can be used as therapeutics for the treatment and/or prevention of cancer.

**[0326]** Because SGA-56M and SGA-56Mv are generally expressed as intracellular proteins, it is preferred that internalizing antibodies be used. However, lipofectin or liposomes can be used to deliver an SGA-56M and/or SGA-56Mv antibody or a fragment of an Fab region thereof into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to an SGA-56M and/or SGA-56Mv polypeptide is preferred. For example, peptides having an amino acid sequence corresponding to the domain of the variable region of an SGA-56M and/or SGA-56Mv specific antibody can be used. Such peptides can be synthesized chemically or produced via recombinant DNA technology using methods well known in the art (*e.g.*, *see* Creighton, 1983, *supra*; and Sambrook *et al.*, 1989, *supra*). Alternatively, single chain antibodies, such as neutralizing antibodies, can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population by utilizing, for example, techniques such as those described in Marasco *et al.* (Marasco,. *et al.*, 1993, *Proc. Natl. Acad. Sci. USA* 90:7889).

**[0327]** The invention also contemplates methods wherein SGA-56M and/or SGA-56Mv antibodies conjugated to a cytostatic and/or a cytotoxic agent are used for treating a patient with a cancer. A useful class of cytotoxic or cytostatic agents which may be conjugated to an antibody of the invention, include, but are not limited to, the following non-mutually exclusive classes of agents: alkylating agents, anthracyclines, antibiotics, antifolates, antimetabolites, antitubulin agents, auristatins, chemotherapy sensitizers, DNA minor groove binders, DNA replication inhibitors, duocarmycins, etoposides, fluorinated pyrimidines, lexitropsins, nitrosoureas, platinols, purine antimetabolites, puromycins, radiation sensitizers, steroids, taxanes, topoisomerase inhibitors, and vinca alkaloids.

**[0328]** Individual cytotoxic or cytostatic agents encompassed by the invention include but are not limited to an androgen, anthramycin (AMC), asparaginase, 5-azacytidine, azathioprine, bleomycin, busulfan, buthionine sulfoximine, camptothecin, carboplatin, carmustine (BSNU), CC-1065, chlorambucil, cisplatin, colchicine, cyclophosphamide, cytarabine, cytidine arabinoside, cytochalasin B, dacarbazine, dactinomycin (formerly actinomycin), daunorubicin, decarbazine, docetaxel, doxorubicin, estrogen, 5-fluordeoxyuridine, 5-fluorouracil, gramicidin D, hydroxyurea, idarubicin, ifosfamide, irinotecan, lomustine (CCNU), mechlorethamine, melphalan, 6-mercaptopurine, methotrexate, mithramycin, mitomycin C, mitoxantrone,

nitroimidazole, paclitaxel, plicamycin, procarbizine, streptozotocin, tenoposide, 6-thioguanine, thioTEPA, topotecan, vinblastine, vincristine, vinorelbine, VP-16 and VM-26.

**[0329]** In a preferred embodiment, a cytotoxic or cytostatic agent is an antimetabolite. The antimetabolite can be a purine antagonist (*e.g.*, azothioprine or mycophenolate mofetil), a dihydrofolate reductase inhibitor (*e.g.*, methotrexate), acyclovir, gangcyclovir, zidovudine, vidarabine, ribavarin, azidothymidine, cytidine arabinoside, amantadine, dideoxyuridine, iododeoxyuridine, poscarnet, and trifluridine.

**[0330]** Techniques for conjugating such therapeutic moieties to proteins, and in particular to antibodies, are well known, *see, e.g.*, Arnon *et al.*, "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies And Cancer Therapy, Reisfeld *et al.* (eds.), pp. 243-56 (Alan R. Liss, Inc., 1985); Hellstrom *et al.*, "Antibodies For Drug Delivery", in Controlled Drug Delivery (2nd ed.), Robinson *et al.* (eds.), pp. 623-53 (Marcel Dekker, Inc., 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in Monoclonal Antibodies '84: Biological And Clinical Applications, Pinchera *et al.* (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in Monoclonal Antibodies For Cancer Detection And Therapy, Baldwin *et al.* (eds.), pp. 303-16 (Academic Press 1985), and Thorpe *et al.*, 1982, Immunol. Rev. 62:119-58.

### 5.6.4. TARGETED DISRUPTION OF SGA-56M AND/OR SGA-56Mv EXPRESSION

**[0331]** As briefly described in Section 5.2.4, *supra*, endogenous SGA-56M and/or SGA-56Mv gene expression can also be reduced by inactivating or "knocking out" the gene or its promoter using targeted homologous recombination. (*e.g.*, *see* Smithies *et al.*, 1985, *Nature* 317:230; Thomas & Capecchi, 1987, *Cell* 51:503; Thompson *et al.*, 1989 *Cell* 5:313). For example, a mutant, non-functional SGA-56M and/or SGA-56Mv gene flanked by DNA homologous to the endogenous SGA-56M and/or SGA-56Mv gene (either the coding regions or regulatory regions of the SGA-56M and/or SGA-56Mv gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express SGA-56M and/or SGA-56Mv *in vivo*. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the SGA-56M and/or SGA-56Mv gene. Such approaches are particularly suited where modifications to ES (embryonic stem) cells can be used to generate animal offspring with an inactive SGA-56Mand/or SGA-56Mv gene homolog (*e.g.*, *see* Thomas & Capecchi 1987 *supra* and Thompson 1989, *supra*). Such techniques can also be utilized to generate animal models of breast cancer and other types of cancer. It should be noted that this approach can be adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate vectors, *e.g.*, herpes virus vectors, retrovirus vectors, adenovirus vectors, or adeno associated virus vectors.

**[0332]** Alternatively, endogenous SGA-56M and/or SGA-56Mv gene expression can be reduced by targeting deoxyribonucleotide sequences complementary to the regulatory region of the SGA-56M and/or SGA-56Mv genes (*i.e.*, the SGA-56M and/or SGA-56Mv gene promoter and/or enhancers) to form triple helical structures that prevent transcription of the SGA-56M and/or SGA-56Mv gene in target cells in the body. (*See generally,*

Helene, 1991, *Anticancer Drug Des.* 6(6):569; Helene *et al.*, 1992, *Ann, N.Y. Acad. Sci.* 660:27; and Maher, 1992, *Bioassays* 14(12):807).

### 5.6.5. COMBINATION THERAPIES

**[0333]** The administration of an SGA-56Mand/or SGA-56Mv antagonist may be used in conjunction with an anti-cancer agent to potentiate the effect of either or both the anti-cancer agent(s) and the antagonist. In a preferred embodiment, the invention further encompasses the use of combination therapy to prevent or treat cancer. In one embodiment, an SGA-56M and/or SGA-56Mv antagonist antagonizes (i.e., reduces or inhibits) SGA-56M and/or SGA-56Mv expression or activity. In yet another embodiment, the SGA-56M and/or SGA-56Mv antagonist reduces or inhibits either SGA-56M and/or SGA-56Mv expression or activity.

**[0334]** In one embodiment, breast cancer and other cancers (*e.g.*, ovarian, lymphoid or skin cancer) can be treated with a pharmaceutical composition comprising an SGA-56M and/or SGA-56Mv antagonist in combination with, for example, 5-fluorouracil, cisplatin, docetaxel, doxorubicin, Herceptin®, gemcitabine (Seidman, 2001, Oncology 15:11-14), IL-2, paclitaxel, and/or VP-16 (etoposide).

**[0335]** Such combination therapies may also be used to prevent cancer, prevent the recurrence of cancer, or prevent the spread or metastasis of a cancer.

**[0336]** Combination therapy also includes, in addition to administration of an SGA-56M and/or SGA-56Mv antagonist, the use of one or more molecules, compounds or treatments that aid in the prevention or treatment of cancer (*i.e.*, cancer therapeutics), which molecules, compounds or treatments includes, but is not limited to, chemoagents, immunotherapeutics, cancer vaccines, anti-angiogenic agents, cytokines, hormone therapies, gene therapies, and radiotherapies.

**[0337]** In one embodiment, one or more chemoagents, in addition to an SGA-56M and/or SGA-56Mv antagonist, is administered to treat a cancer patient. A chemoagent (or "anti-cancer agent" or "anti-tumor agent" or "cancer therapeutic") refers to any molecule or compound that assists in the treatment of a cancer. Examples of chemoagents contemplated by the present invention include, but are not limited to, cytosine arabinoside, taxoids (*e.g.*, paclitaxel, docetaxel), anti-tubulin agents (*e.g.*, paclitaxel, docetaxel, epothilone B, or its analogues), macrolides (*e.g.*, rhizoxin) cisplatin, carboplatin, adriamycin, tenoposide, mitozantron, discodermolide, eleutherobine, 2-chlorodeoxyadenosine, alkylating agents (*e.g.*, cyclophosphamide, mechlorethamine, thioepa, chlorambucil, melphalan, carmustine (BSNU), lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin, thio-tepa), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, anthramycin), antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, flavopiridol, 5-fluorouracil, fludarabine, gemcitabine, dacarbazine, temozolamide), asparaginase, *Bacillus Calmette and Guerin*, diphtheria toxin, hexamethylmelamine, hydroxyurea, LYSODREN®, nucleoside analogues, plant alkaloids (*e.g.*, Taxol, paclitaxel, camptothecin, topotecan, irinotecan (CAMPTOSAR, CPT-11), vincristine, vinca alkyloids such as vinblastine), podophyllotoxin (including derivatives such as epipodophyllotoxin, VP-16 (etoposide), VM-26 (teniposide)), cytochalasin B, colchine, gramicidin D,

ethidium bromide, emetine, mitomycin, procarbazine, mechlorethamine, anthracyclines (*e.g.*, daunorubicin (formerly daunomycin), doxorubicin, doxorubicin liposomal), dihydroxyanthracindione, mitoxantrone, mithramycin, actinomycin D, procaine, tetracaine, lidocaine, propranolol, puromycin, anti-mitotic agents, abrin, ricin A, pseudomonas exotoxin, nerve growth factor, platelet derived growth factor, tissue plasminogen activator, aldesleukin, allutamine, anastrozle, bicalutamide, biaomycin, busulfan, capecitabine, carboplain, chlorabusil, cladribine, cylarabine, daclinomycin, estramusine, floxuridhe, gamcitabine, gosereine, idarubicin, itosfamide, lauprolide acetate, levamisole, lomusline, mechlorethamine, magestrol, acetate, mercaptopurino, mesna, mitolanc, pegaspergase, pentoslatin, picamycin, riuxlmab, campath-1, straplozocin, thioguanine, tretinoin, vinorelbine, or any fragments, family members, or derivatives thereof, including pharmaceutically acceptable salts thereof. Compositions comprising one or more chemoagents (*e.g.*, FLAG, CHOP) are also contemplated by the present invention. FLAG comprises fludarabine, cytosine arabinoside (Ara-C) and G-CSF. CHOP comprises cyclophosphamide, vincristine, doxorubicin, and prednisone.

**[0338]** In one embodiment, said chemoagent is gemcitabine at a dose ranging from 100 to 1000 mg/m$^2$/cycle. In one embodiment, said chemoagent is dacarbazine at a dose ranging from 200 to 4000 mg/m$^2$/cycle. In a preferred embodiment, said dose ranges from 700 to 1000 mg/m$^2$/cycle. In another embodiment, said chemoagent is fludarabine at a dose ranging from 25 to 50 mg/m$^2$/cycle. In another embodiment, said chemoagent is cytosine arabinoside (Ara-C) at a dose ranging from 200 to 2000 mg/m$^2$/cycle. In another embodiment, said chemoagent is docetaxel at a dose ranging from 1.5 to 7.5 mg/kg/cycle. In another embodiment, said chemoagent is paclitaxel at a dose ranging from 5 to 15 mg/kg/cycle. In yet another embodiment, said chemoagent is cisplatin at a dose ranging from 5 to 20 mg/kg/cycle. In yet another embodiment, said chemoagent is 5-fluorouracil at a dose ranging from 5 to 20 mg/kg/cycle. In yet another embodiment, said chemoagent is doxorubicin at a dose ranging from 2 to 8 mg/kg/cycle. In yet another embodiment, said chemoagent is epipodophyllotoxin at a dose ranging from 40 to 160 mg/kg/cycle. In yet another embodiment, said chemoagent is cyclophosphamide at a dose ranging from 50 to 200 mg/kg/cycle. In yet another embodiment, said chemoagent is irinotecan at a dose ranging from 50 to 150 mg/m$^2$/cycle. In yet another embodiment, said chemoagent is vinblastine at a dose ranging from 3.7 to 18.5 mg/m$^2$/cycle. In yet another embodiment, said chemoagent is vincristine at a dose ranging from 0.7 to 2 mg/m$^2$/cycle. In yet another embodiment, said chemoagent is methotrexate at a dose ranging from 3.3 to 1000 mg/m$^2$/cycle.

**[0339]** In a preferred embodiment, the invention further encompasses the use of low doses of chemoagents when administered in conjunction with an SGA-56M and/or SGA-56Mv antagonist in a combination treatment regimen. For example, initial treatment with an SGA-56M and/or SGA-56Mv antagonist increases the sensitivity of a tumor to subsequent challenge with a dose of chemoagent, which dose is near or below the lower range of dosages when the chemoagent is administered in the absence of an SGA-56M and/or SGA-56Mv antagonist. In one embodiment, an SGA-56M and/or SGA-56Mv antagonist and a low dose (*e.g.*, 6 to 60 mg/m$^2$/day or less) of docetaxel are administered to a cancer patient. In another embodiment, an SGA-56M and/or SGA-56Mv antagonist and a low dose (*e.g.*, 10 to 135 mg/m$^2$/day or less) of paclitaxel are administered to a cancer patient. In yet another embodiment, an SGA-56M and/or SGA-56Mv antagonist and

a low dose (*e.g.*, 2.5 to 25 mg/m$^2$/day or less) of fludarabine are administered to a cancer patient. In yet another embodiment, an SGA-56M and/or SGA-56Mv antagonist and a low dose (*e.g.*, 0.5 to 1.5 g/m$^2$/day or less) of cytosine arabinoside (Ara-C) are administered to a cancer patient.

**[0340]** The invention, therefore, contemplates the use of one or more SGA-56M and/or SGA-56Mv antagonists, which is administered prior to, subsequently, or concurrently with low doses of chemoagents, for the prevention or treatment of cancer.

**[0341]** In one embodiment, said chemoagent is gemcitabine at a dose ranging from 10 to 100mg/m$^2$/cycle.

**[0342]** In one embodiment, said chemoagent is cisplatin, *e.g.*, PLATINOL™ or PLATINOL-AQ™(Bristol Myers), at a dose ranging from 5 to 75 mg/m$^2$/cycle. In another embodiment, a dose of cisplatin ranging from 7.5 to 75 mg/m$^2$/cycle is administered to a patient with ovarian cancer or other cancer. In another embodiment, a dose of cisplatin ranging from 5 to 50 mg/m$^2$/cycle is administered to a patient with bladder cancer or other cancer.

**[0343]** In another embodiment, said chemoagent is carboplatin, *e.g.*, PARAPLATIN™(Bristol Myers), at a dose ranging from 2 to 75 mg/m$^2$/cycle. In another embodiment, a dose of carboplatin ranging from 7.5 to 75 mg/m$^2$/cycle is administered to a patient with ovarian cancer or other cancer. In another embodiment, a dose of carboplatin ranging from 5 to 50 mg/m$^2$/cycle is administered to a patient with bladder cancer or other cancer. In another embodiment, a dose of carboplatin ranging from 2 to 20 mg/m$^2$/cycle is administered to a patient with testicular cancer or other cancer.

**[0344]** In another embodiment, said chemoagent is docetaxel, *e.g.*, TAXOTERE™ (Rhone Poulenc Rorer) at a dose ranging from 6 to 60 mg/m$^2$/cycle.

**[0345]** In another embodiment, said chemoagent is paclitaxel, *e.g.*, TAXOL™ (Bristol Myers Squibb), at a dose ranging from 10 to 135 mg/kg/cycle.

**[0346]** In another embodiment, said chemoagent is 5-fluorouracil at a dose ranging from 0.5 to 5 mg/kg/cycle.

**[0347]** In another embodiment, said chemoagent is doxorubicin, *e.g.*, ADRIAMYCIN™ (Pharmacia & Upjohn), DOXIL (Alza), RUBEX™ (Bristol Myers Squibb), at a dose ranging from 2 to 60 mg/kg/cycle.

**[0348]** In another embodiment, an SGA-56M and/or SGA-56Mv antagonist is administered in combination with one or more immunotherapeutic agents, such as antibodies or immunomodulators, which include, but are not limited to, Herceptin®, Retuxan®, OvaRex, Panorex, BEC2, IMC-C225, Vitaxin, Campath I/H, Smart MI95, LymphoCide, Smart I D10, and Oncolym, rituxan, rituximab, gemtuzumab, or trastuzumab.

**[0349]** In another embodiment, an SGA-56M and/or SGA-56Mv antagonist is administered in combination with one or more anti-angiogenic agents, which includes, but is not limited to, angiostatin, thalidomide, kringle 5, endostatin, Serpin (Serine Protease Inhibitor) anti-thrombin, 29 kDa N–terminal and a 40 kDa C-terminal proteolytic fragments of fibronectin, 16 kDa proteolytic fragment of prolactin, 7.8 kDa proteolytic fragment of platelet factor-4 , a 13-amino acid peptide corresponding to a fragment of platelet factor-4 (Maione et al., 1990, *Cancer Res.* 51:2077), a 14-amino acid peptide corresponding to a fragment of collagen I (Tolma et al., 1993, *J. Cell Biol.* 122:497), a 19 amino acid peptide corresponding to a fragment of

Thrombospondin I (Tolsma et al., 1993, *J. Cell Biol.* 122:497), a 20-amino acid peptide corresponding to a fragment of SPARC (Sage et al., 1995, *J. Cell. Biochem.* 57:1329-), or any fragments, family members, or derivatives thereof, including pharmaceutically acceptable salts thereof.

**[0350]** Other peptides that inhibit angiogenesis and correspond to fragments of laminin, fibronectin, procollagen, and EGF have also been described (*See* the review by Cao, 1998, *Prog. Mol. Subcell. Biol.* 20:161). Monoclonal antibodies and cyclic pentapeptides, which block certain integrins that bind RGD proteins (*i.e.*, possess the peptide motif Arg-Gly-Asp), have been demonstrated to have anti-vascularization activities (Brooks *et al.*, 1994, *Science* 264:569; Hammes *et al.*, 1996, *Nature Medicine* 2:529). Moreover, inhibition of the urokinase plasminogen activator receptor by antagonists inhibits angiogenesis, tumor growth and metastasis (Min *et al.*, 1996, *Cancer Res.* 56:2428-33; Crowley *et al.*, 1993, *Proc Natl Acad Sci.USA* 90:5021). Use of such anti-angiogenic agents is also contemplated by the present invention.

**[0351]** In another embodiment, an SGA-56M and/or SGA-56Mv antagonist is administered in combination with a regimen of radiation.

**[0352]** In another embodiment, an SGA-56M and/or SGA-56Mv antagonist is administered in combination with one or more cytokines, which includes, but is not limited to, lymphokines, tumor necrosis factors, tumor necrosis factor-like cytokines, lymphotoxin-α, lymphotoxin-β, interferon-α, interferon-β, macrophage inflammatory proteins, granulocyte monocyte colony stimulating factor, interleukins (including, but not limited to, interleukin-1, interleukin-2, interleukin-6, interleukin-12, interleukin-15, interleukin-18), OX40, CD27, CD30, CD40 or CD137 ligands, Fas-Fas ligand, 4-1BBL, endothelial monocyte activating protein or any fragments, family members, or derivatives thereof, including pharmaceutically acceptable salts thereof.

**[0353]** In yet another embodiment, an SGA-56M and/or SGA-56Mv antagonist is administered in combination with a cancer vaccine. Examples of cancer vaccines include, but are not limited to, autologous cells or tissues, non-autologous cells or tissues, carcinoembryonic antigen, alpha-fetoprotein, human chorionic gonadotropin, BCG live vaccine, melanocyte lineage proteins (*e.g.*, gp100, MART-1/MelanA, TRP-1 (gp75), tyrosinase, widely shared tumor-associated, including tumor-specific, antigens (*e.g.*, BAGE, GAGE-1, GAGE-2, MAGE-1, MAGE-3, N-acetylglucosaminyltransferase-V, p15), mutated antigens that are tumor-associated (β-catenin, MUM-1, CDK4), nonmelanoma antigens (*e.g.*, HER-2/neu (breast and ovarian carcinoma), human papillomavirus-E6, E7 (cervical carcinoma), MUC-1 (breast, ovarian and pancreatic carcinoma). For human tumor antigens recognized by T-cells, *see generally* Robbins and Kawakami, 1996, *Curr. Opin. Immunol.* 8:628. Cancer vaccines may or may not be purified preparations.

**[0354]** In yet another embodiment, an SGA-56M and/or SGA-56Mv antagonist is used in association with a hormonal treatment. Hormonal therapeutic treatments comprise hormonal agonists, hormonal antagonists (*e.g.*, flutamide, tamoxifen, leuprolide acetate (LUPRON), LH-RH antagonists), inhibitors of hormone biosynthesis and processing, and steroids (*e.g.*, dexamethasone, retinoids, betamethasone, cortisol, cortisone, prednisone, dehydrotestosterone, glucocorticoids, mineralocorticoids, estrogen, testosterone, progestins), antigestagens (*e.g.*, mifepristone, onapristone), and antiandrogens (*e.g.*, cyproterone acetate).

**[0355]** In yet another embodiment, an SGA-56M and/or SGA-56Mv antagonist is used in association with a gene therapy program in the treatment of cancer. In one embodiment, gene therapy with recombinant cells secreting interleukin-2 is administered in combination with an SGA-56M and/or SGA-56Mv antagonist to prevent or treat cancer, particularly breast cancer (*See, e.g.*, Deshmukh *et al.*, 2001, *J. Neurosurg.* 94:287).

**[0356]** In one embodiment, an SGA-56M and/or SGA-56Mv antagonist is administered, in combination with at least one cancer therapeutic agent, to a cancer patient for a short treatment cycle to ameliorate the symptoms of the cancer and potentially eliminate the cancer. The duration of treatment with the cancer therapeutic agent may vary according to the particular cancer therapeutic agent used. The invention also contemplates discontinuous administration or daily doses divided into several partial administrations. Appropriate treatment time-lines for cancer therapeutic agents will be appreciated by those skilled in the art, and the invention contemplates the continued assessment of optimal treatment schedules for each cancer therapeutic agent.

**[0357]** The present invention contemplates at least one cycle, preferably more than one cycle during which a single therapeutic or sequence of therapeutics is administered. An appropriate period of time for one cycle will be appreciated by the skilled artisan, as will the total number of cycles, and the interval between cycles. The invention contemplates the continued assessment of optimal treatment schedules for each SGA-56M and/or SGA-56Mv antagonist and cancer therapeutic agent.

## 5.7. PHARMACEUTICAL PREPARATIONS AND METHODS OF ADMINISTRATION

**[0358]** The compounds, proteins, peptides, nucleic acid sequences and fragments thereof, described herein can be administered to a patient at therapeutically effective doses to treat cancer, *e.g.*, breast cancer wherein the expression level of the SGA-56M and/or SGA-56Mv gene is elevated compared to a non-cancerous sample or a predetermined non-cancerous standard. A therapeutically effective dose refers to that amount of a compound sufficient to result in a healthful benefit in the treated subject.

### 5.7.1. EFFECTIVE DOSE

**[0359]** Toxicity and therapeutic efficacy of compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the $LD_{50}$ (the dose lethal to 50% of the population) and the $ED_{50}$ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio $LD_{50}/ED_{50}$. Compounds that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to unaffected cells and, thereby, reduce side effects.

**[0360]** The data obtained from cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the $ED_{50}$ with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used

in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the $IC_{50}$ (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured by any technique known in the art, for example, by high performance liquid chromatography.

### 5.7.2. FORMULATIONS AND USE

**[0361]** The invention relates to pharmaceutical compositions, including, but not limited to pharmaceutical compositions comprising an SGA-56M and/or SGA-56Mv gene product, or antagonists thereof, for the treatment or prevention of cancer.

**[0362]** Pharmaceutical compositions for use in accordance with the present invention, *e.g.*, methods to treat or prevent cancer, can be formulated in a conventional manner using one or more physiologically acceptable carriers or excipients.

**[0363]** Thus, the compounds and their physiologically acceptable salts and solvents can be formulated for administration by inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral or rectal administration.

**[0364]** For oral administration, the pharmaceutical compositions can take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (*e.g.*, pregelatinized maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (*e.g.*, lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (*e.g.*, magnesium stearate, talc or silica); disintegrants (*e.g.*, potato starch or sodium starch glycolate); or wetting agents (*e.g.*, sodium lauryl sulphate). The tablets can be coated by methods well known in the art. Liquid preparations for oral administration can take the form of, for example, solutions, syrups or suspensions, or they can be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations can be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (*e.g.*, sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (*e.g.*, lecithin or acacia); non-aqueous vehicles (*e.g.*, almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (*e.g.*, methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations can also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

**[0365]** Preparations for oral administration can be suitably formulated to give controlled release of the active compound.

**[0366]** For buccal administration the compositions can take the form of tablets or lozenges formulated in conventional manner.

**[0367]** For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane,

dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit can be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, e.g., gelatin for use in an inhaler or insufflator can be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

**[0368]** The compounds can be formulated for parenteral administration (*i.e.*, intravenous or intramuscular) by injection, via, for example, bolus injection or continuous infusion. Formulations for injection can be presented in unit dosage form, *e.g.*, in ampoules or in multi-dose containers, with an added preservative. The compositions can take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and can contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient can be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

**[0369]** The compounds can also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides.

**[0370]** In addition to the formulations described previously, the compounds can also be formulated as a depot preparation. Such long acting formulations can be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds can be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

## 5.8. VACCINE THERAPY

**[0371]** Peptides and proteins encoded by the SGA-56M or SGA-56Mv gene and nucleic acids which encode an SGA-56M or SGA-56Mv polypeptide or fragments thereof, can be used as vaccines by administering to an individual at risk of developing cancer an amount of said protein, peptide, or nucleic acid that effectively stimulates an immune response against an SGA-56M or SGA-56Mv -encoded polypeptide and protects that individual from cancer. The invention thus contemplates a method of vaccinating a subject against cancer wherein said subject at risk for developing a cancer.

**[0372]** Many methods may be used to introduce the vaccine formulations described above, these include but are not limited to intranasal, intratracheal, oral, intradermal, intramuscular, intraperitoneal, intravenous, and subcutaneous routes. Various adjuvants may be used to increase the immunological response, and include but are not limited to, Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and corynebacterium parvum. Such adjuvants are also well known in the art.

[0373] The nucleotides of the invention, including variants and derivatives, can be used as vaccines, *e.g.*, by genetic immunization. Genetic immunization is particularly advantageous as it stimulates a cytotoxic T-cell response but does not utilize live attenuated vaccines, which can revert to a virulent form and infect the host causing complications from infection. As used herein, genetic immunization comprises inserting the nucleotides of the invention into a host, and that the nucleotide uptake by the host cells and the proteins encoded by the nucleotides are translated. These translated proteins are then either secreted or processed by the host cell for presentation to immune cells and an immune reaction is stimulated. Preferably, the immune reaction is a cytotoxic T cell response, however, a humoral response or macrophage stimulation is also useful in preventing initial or additional tumor growth and metastasis or spread of a cancer. The skilled artisan will appreciate that there are various methods for introducing foreign nucleotides into a host animal and subsequently into cells for genetic immunization, for example, by intramuscular injection of about 50 mg of plasmid DNA encoding the proteins of the invention solubilized in 50 ml of sterile saline solution, with a suitable adjuvant (*See, e.g.*, Weiner and Kennedy, 1999, *Scientific American* 7:50-57; Lowrie et al., 1999, *Nature* 400:269-271).

[0374] The invention thus provides a vaccine formulation for the prevention and/or treatment of cancer comprising an immunogenic amount of an SGA-56M or SGA-56Mv gene product. The invention further provides for an immunogenic composition comprising a purified SGA-56M or SGA-56Mv gene product.

## 5.9. KITS

[0375] The invention includes a kit for assessing the presence of cancer cells including breast cancer cells (*e.g.*, in a sample such as a patient sample). The kit comprises a plurality of reagents, each of which is capable of binding specifically with a nucleic acid or polypeptide corresponding to a marker of the invention, *e.g.*, the SGA-56M or SGA-56Mv gene or gene product or fragment thereof. Suitable reagents for binding with a polypeptide corresponding to a marker of the invention include antibodies, antibody derivatives, labeled antibodies, antibody fragments, and the like. Suitable reagents for binding with a nucleic acid (*e.g.*, a genomic DNA, an mRNA, a spliced mRNA, a cDNA, or the like) include complementary nucleic acids. For example, the nucleic acid reagents may include oligonucleotides (labeled or non-labeled) fixed to a substrate, labeled oligonucleotides not bound with a substrate, pairs of PCR primers, molecular beacon probes, and the like.

[0376] The kit of the invention may optionally comprise additional components useful for performing the methods of the invention. By way of example, the kit may comprise fluids (*e.g.*, SSC buffer) suitable for annealing complementary nucleic acids or for binding an antibody with a protein with which it specifically binds, one or more sample compartments, an instructional material which describes performance of a method of the invention, a sample of normal cells, a sample of cancer cells, and the like.

## 6. EXAMPLES

[0377] The isolation of novel breast cancer-associated antigens SGA-56M and SGA-56Mv (Seattle Genetics Antigen isolated from MCF-7 cells, and variants (v) thereof) is described. The nucleic acid sequences of SGA-56M (SEQ ID NO: 1) and SGA-56Mv (SEQ ID NO: 3) have been deposited with GenBank. MCF-7 is an estrogen receptor positive (ER+) breast adenocarcinoma cell-line. Suppression Subtractive Hybridization (SSH) and high-throughput cDNA microarrays were combined in analyzing genes over-expressed in breast cancer. The results detail the effectiveness of combining SSH and cDNA microarrays in providing breast cancer-specific expression profiles. Sequence analysis revealed heretofore uncharacterized molecules SGA-56M and SGA-56Mv, and several previously identified cancer-specific genes. The SGA-56M cDNA (FIG. 1) and SGA-56Mv cDNA (FIG. 2) were cloned by PCR and sequence verified by automated fluorescent sequencing (Applied Biosystems, Foster City, CA). Based on their tumor selectivity (described in Section 6.3), SGA-56M and SGA-56Mv are useful therapeutic targets and/or diagnostic markers in the treatment of breast cancer, lung cancer, and other SGA-56M or SGA-56Mv positive cancers.

### 6.1. INTRODUCTION

[0378] Breast cancer arises from a malignancy of epithelial cells in the female, and occasionally the male, usually of adenocarcinomal origin initiated in the ductal breast epithelium. The majority of breast cancer cases are estrogen-dependent adenocarcinomas. The MCF-7 breast cancer-derived tumor cell line is an estrogen-dependent example. Breast Cancer is the most common non-dermal malignancy in women and 192,200 cases are anticipated in the U.S. for the upcoming year. Despite recent advances in early diagnosis and treatment, 40,200 U.S. women have succumbed to this disease in the year 2000 (Greenlee *et al.*, 2001, *Cancer Statistics* 51(1): 15). Breast cancer, second only to lung cancer in mortality rates annually, requires continued discovery of additional uncharacterized antigens and innovative utility of these molecules to improve overall therapy and intervention.

[0379] In total, 10% of all breast cancers are initiated by a genetic mutation similar to BRCA-1 and BRCA-2 (Nathanson *et al.*, 2001, *Nature Med.* 7(5): 552). The transformation of normal epithelium and progression to metastatic breast cancer arises from a cascade of genetic alterations that translate to global changes in cellular protein composition and expression. Some of these changes, detected in the form of cell-surface markers, are the focus of present diagnostic and tumor targeting efforts. For example, the HER-2/neu oncogene, which encodes a 185-kDa protein transmembrane protein, is overexpressed in 10-30% of invasive breast cancers, 40-60% of intraductal breast carcinomas, as well as other cancer types (Koeppen *et al.*, 2001, *Histopathology* 38(2): 96). Antibodies to HER2-neu (Herceptin®) have been shown to identify and selectively sensitize antigen positive cells to anti-cancer therapy (Baselga *et al.*, 1998, *Cancer Res.* 58:2825).

[0380] The sex steroid estrogen has been shown to play a major role in tissue development as well as other physiological processes. In addition, it has been reported to play a critical role in the progression of both breast and gynecological cancers (Pike *et al.*, 1993, *Epidemiol. Rev.*15: 17). MCF-7 is a well-established tumor cell-line that is an ER+ adenocarcinoma. Despite its existence in cell-culture for nearly three decades,

it remains likely that many durable alterations in gene expression patterns still persist since its isolation and initial characterization in 1973 (Brooks *et al.*, 1973, *J. Biol. Chem.* 248(17): 6251). Some of the stable genes, and specifically SGA-56M or SGA-56Mv as described herein provide potential targets for diagnostic or therapeutic strategies for breast cancer.

**[0381]** To evaluate the hypothesis that many of such targets have remained unrecognized, tumor-enriched SSH libraries were constructed and arrayed to selectively screen for tumor-specific genes. SSH is a technique well known in the art for its effectiveness in characterizing and prioritizing differentially expressed genes: (Chu *et al.*, 1997, *Proc. Natl. Acad. Sci.* 94(19): 10057; Gurskaya *et al.*, 1996, *Anal. Biochem.* 240: 90; Kuang *et al.*,1998, *Nuc. Acid Res.* 26: 1116; von Stein *et al.*, 1997, *Nuc. Acid Res.* 25: 2598; Wong *et al.*, 1997, *J. Biol. Chem.* 272(40): 25190; and Yokomizo *et al.*, 1997, *Nature* 387: 620). SGA-56M or SGA-56Mv, novel breast cancer-associated proteins, were discovered utilizing these techniques as described herein. The initial tumor-enriched MCF-7-specific SSH libraries were evaluated in a higher density format with minimal redundancy, demonstrating that the overall complexity of the libraries had not been compromised.

**[0382]** Intensive and systematic evaluation of gene expression patterns is crucial in understanding the physiological mechanisms associated with cellular transformation and metastasis. Currently, several technical platforms are being used to accomplish this goal. They include: Serial Analysis of Gene Expression (SAGE) (Velculescu *et al.*, 1995, *Science* 270: 484), Restriction Enzyme Analysis of Differentially Expressed Sequences (READS) (Prasher *et al.*, 1999, *Methods Enzymol.* 303: 258), Amplified Fragment Length Polymorphism (AFLP) (Bachem *et al.*, 1996, *Plant J.* 9: 745), Representational Difference Analysis (RDA) (Hubank *et al.*, 1994, *Nucleic Acid Res.* 22(25): 5640), Differential Display (Liang *et al.*, 1992, *Cancer Res.* 52(24): 6966) and SSH (Diatchenko *et al.*, 1996, *Proc. Natl. Acad. Sci.* 93: 6025) as detailed in this text. SSH is very similar to RDA with the exception of an additional normalization step that is included to increase the relative abundance of rare transcripts. The combination of SSH and cDNA microarrays offers several advantages over the aforementioned technologies in the discovery of novel tumor-associated proteins and antigens (TAA's). The use of SSH is an attractive approach to identifying novel cancer targets because it does not rely on previously characterized cDNA sets. SSH efficiently normalizes both frequent and rare transcripts at equivalent levels and preferentially amplifies only those which are differentially expressed. The use of expression arrays further increases the chances of identifying lead targets by examining thousands of genes in a single experiment.

## 6.2. MATERIALS AND METHODS

### 6.2.1. CELL CULTURE

**[0383]** Breast tumor cell-lines MCF-7, T47-D, SKBR-3, MDA-MB-231, MDA-MB-435s, MDA-MB-453, H3396, Hs578T and BT-549 were grown in RPMI 1640 medium® supplemented with 10% fetal bovine serum plus 100 U/mL penicillin G and 100 µg/mL streptomycin sulfate. All tumor cell-lines were passaged once per week by trypsinization and replated at 2500-3000 cells/$cm^2$. Normal human mammary epithelial

cells (HMEC) were maintained in MEGM® (Clonetics, San Diego, CA). HMEC's were passaged once per week by trypsinization and replating at 2500-3000 cells/$cm^2$.

### 6.2.2. RNA ISOLATION

**[0384]** Total RNA was isolated from cultured cells using RNA-Bee™ (Tel-Test, Inc., Friendswood, TX). Poly A+ RNA was extracted using the Oligotex mRNA Midi kit ® (Qiagen, Inc., Valencia, CA).

### 6.2.3. GENERATION OF SSH cDNA LIBRARIES

**[0385]** MCF-7 breast cancer-specific SSH cDNA libraries were constructed essentially as described by Diatchenko *et al.*, 1996, *Proc. Natl. Acad. Sci.* 93:6025. Library one was constructed using the breast tumor ER+ cell-line MCF-7 (tester) vs. HMEC (driver). Library two was constructed using the breast tumor ER+ cell-line MCF-7 (tester) vs. a pool of 5 ER- cell lines (SKBR-3, MDA-MB-231, MDA-MB-435s, Hs578T, and BT-549) (driver).

**[0386]** Driver cDNA was synthesized from 2 ug of poly A+ RNA using 1 ul of 10 uM cDNA synthesis primer 5'-$TTTTGTACAAGCTT_{30}N_1N$-3' (SEQ ID NO: 7) and 1 ul of 200 u/ul Superscript II Reverse Transcriptase® (Invitrogen, Carlsbad, CA). The resulting cDNA pellet was digested with 1.5 ul of 10u/ul of *Rsa* I restriction enzyme. Driver cDNA's were then precipitated with 100 ul of 10M Ammonium Acetate (Sigma, St. Louis, MO), 3 ul of 20 mg/ml glycogen (Roche Molecular Biochemicals, Indianapolis, IN) and 1 ml of ethanol (Sigma, St. Louis, MO). The cDNA preparations were then resuspended in 5 ul of diethyl pyrocarbonate (DEPC) treated water.

**[0387]** Tester cDNA was synthesized from 2 ug of poly A+ RNA as described above for the driver. Rsa I digested tester cDNA was diluted in 5 ul of DEPC treated water prior to adaptor ligation. Diluted tester cDNA (2 ul) was ligated to 2 ul of 10 uM adaptor 1 (5'-CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT-3') (SEQ ID NO: 8) and 2 ul of 10 uM adaptor 2R (5'-CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGT-3') (SEQ ID NO: 9) in separate reactions using 0.5 units of T4 DNA ligase (Invitrogen, Carlsbad, CA).

**[0388]** Driver cDNA (600 ng) was added separately to each of the two tubes containing adaptor-1 ligated tester (20 ng) and adaptor 2R ligated tester (20 ng). The samples were mixed, ethanol precipitated as described above, and resuspended in 1.5 ul of hybridization buffer (50 mM Hepes pH 8.3, 0.5 M NaCl/0.0.2 mM EDTA pH 8.0). The reaction mixture was placed in hot start PCR tubes, (Molecular BioProducts, San Diego, CA), denatured at 95°C for 1.5 min. and then incubated at 68°C for 8 hrs. After this initial hybridization, the samples were combined and excess heat denatured driver cDNA (150 ng) was added. This secondary reaction mixture was incubated overnight at 68°C. The final hybridization mixture was diluted in 200 ul of dilution buffer (20 mM Hepes pH 8.3, 50mM NaCl, 0.2 mM EDTA) and stored at -20°C.

**[0389]** Two rounds of PCR amplification were performed for each SSH library. The primary PCR was performed in 25 ul. The reaction mixture contained 1 ul of diluted subtracted cDNA, 1 ul of 10 uM PCR primer 1 (5'-CTAATACGACTCACTATAGGGC-3') (SEQ ID NO: 10), 10x PCR buffer consisting of (166

mM $NH_4C_2H_3O_2$, 670 mM Tris pH 8.8, 67 mM $MgCl_2$, and 100mM 2-Mercaptoethanol), 1.5 ul of 10mM dNTP's, 1.5 ul Dimethyl Sulfoxide (DMSO) (Sigma, St. Louis, MO), and 0.25 ul of 5 u/ul of Taq polymerase (Brinkmann, Westbury, NY). PCR was performed with the following cycling conditions: 75°C for 7 min.; 94°C for 2 min.; 27 cycles at 94°C for 30 sec., 66°C for 30 sec., and 72°C for 1.5 min.; and a final extension at 72°C for 5 min. A secondary PCR was performed using 1 ul of the primary PCR as template with the same reaction components as above. Nested PCR primers NP1 (5'-TCGAGCGGCCGCCCGGGCAGGT-3') (SEQ ID NO: 11) and NP2R (5'-AGCGTGGTCGCGGCCGAGGT-3') (SEQ ID NO: 12) were used in place of PCR primer 1. The secondary PCR was performed with the following cycling conditions: 94°C for 2 min.; 15 cycles at 94°C for 30 sec., 68°C for 30 sec., and 72°Cfor 1.5 min.; and a final extension at 72°C for 5 min. The PCR products were analyzed on 1.5% ultrapure agarose gels (Invitrogen, Carlsbad, CA) and visualized by ethidium bromide (Fisher Chemical, Fair Lawn, NJ).

**[0390]** Subtraction efficiency was confirmed by PCR depletion of EF-1 and Tubulin. EF-1 primers were EF-1 (5'-CTGTTCCTGTTGGCCGAGTC-3') (SEQ ID NO: 13) and EF-2 (5' -CGATGCATTGTTATCATTAAC-3') (SEQ ID NO: 14). Tubulin primers were Tu-1 (5'-CACCCTGAGCAGCTCATCAC-3') (SEQ ID NO: 15) and Tu2 (5'-GGCCAGGGTCACATTTCACC-3') (SEQ ID NO: 16).

### 6.2.4. CLONING OF SSH POOLS INTO pCR4-TOPO

**[0391]** The SSH-cDNA pools were cloned into the pCR4-TOPO® vector (Invitrogen, Carlsbad, CA) and transformed into chemically competent TOP 10 cells® (Invitrogen, Carlsbad, CA). The library was plated on LB agar plates (Becton Dickinson, Sparks, MD) containing 50µg/ul kanamycin (Sigma, St. Louis, MO). Cloning efficiency and size distribution for each library was determined by amplification using M13 (-20) (5'-GTAAAACGACGGCCAGT-3') (SEQ ID NO: 17) and Ml 3R (5' -CAGGAAACAGCTATGACC-3') (SEQ ID NO: 18) universal primers.

### 6.2.5. CUSTOM ARRAY GENERATION

**[0392]** SSH clones containing cDNA sequences of interest were amplified using M13 (-20) and M13R universal primers. PCR products were purified using 96-well MultiScreen PCR Purification Plates (Millipore, Bedford, MA). Microarrays were prepared by spotting targets in duplicate on positively charged nylon membranes (Hybond-XL®, Amersham Pharmacia Biotech, Piscataway, NJ) at concentrations of 2 ng DNA/spot using a Biomek 2000 Robot® (Beckman Coulter Inc., Fullerton, CA). For probe construction, mRNA was isolated from cell lines as described above. Poly A+ RNA (1 ug) was converted to cDNA and labeled with ($^{\alpha}$-P32) dCTP (Amersham Pharmacia Biotech, Piscataway, NJ) by reverse transcription using Superscript II RT® (Invitrogen, Carlsbad, CA). Hybridizations were performed overnight at 42°C in 6X Saline Sodium Citrate (SSC), 0.1% Sodium Dodecyl Sulfate (SDS), 50% Deionized Formamide, and 5X Denhardt's solution (1% Ficoll Type 400, 1% polyvinylpyrrolidone, and 1% bovine serum albumin) (Research

Genetics, Huntsville, AL). Wash conditions were 4 times in 2X SSC/0.1% SDS for 10 min. each at room temperature, followed by 4 high stringency washes in 0.1X SSC/0.1%SDS at 65°C for 30 min. each.

### 6.2.6. ARRAY DATA ANALYSIS

**[0393]** Hybridization Intensities were quantitated on the Phosphorlmager SI® (Molecular Dynamics, Sunnyvale, CA) using ArrayVision 5.1 Software® (Imaging Research, St. Catharines, ON, CA). Average signal intensities were determined for each set of duplicate spots. For each membrane analyzed, relative quantitative values were determined based on normalization to multiple housekeeping genes spotted at various locations on each membrane. This allowed for blot-to-blot comparisons in determining differential expression. Two independent microarray experiments were performed for each comparison to ensure overall validity and reproducibility of the results. Targets greater than 2-fold over-expressed in a tumor vs. normal comparison were considered for further evaluation.

### 6.2.7. SEMI-QUANTITATIVE RT-PCR

**[0394]** DNA was synthesized from 5 ug total RNA using the Superscript First-Strand cDNA Synthesis System for RT-PCR® (Invitrogen, Carlsbad, CA). Gene specific primers were selected for SGA-56M or SGA-56Mv and EF-1 to obtain semi-quantitative mRNA levels. Primers used were common for SGA-56M and SGA-56Mv. They were as follows: RT1 (5'-GCTTGGAAAAGTTGAGCC -3') (SEQ ID NO: 19), and RT2 (5'-CTGGGTCTGAGTCTTAGC-3') (SEQ ID NO: 20). Primers for EF-1 were as follows: EF-1 (5'-CTGTTCCTGTTGGCCGAGTC-3') (SEQ ID NO: 13) and EF-2 (5' CGATGCATTGTTATCATTAAC-3') (SEQ ID NO: 14).

### 6.2.8. MULTIPLE TISSUE EXPRESSION ARRAY (MTE™)

**[0395]** The MTE™ (Clontech, Palo Alto, CA) array was used to determine relative expression of SGA-56M/SGA-56Mv in various normal populations. Primers used in amplifying a probe were common for SGA-56M and SGA-56Mv. Primers were as follows: RT1 (5'-GCTTGGAAAAGTTGAGCC -3') (SEQ ID NO: 19), and RT2 (5'-CTGGGTCTGAGTCTTAGC-3') (SEQ ID NO: 20). Fifty ng of PCR product were labeled using Ready-to-go Beads® (Amersham Biosciences Corporation, Piscataway, NJ) and $^{\alpha}$-P32 dCTP at 3000 Ci/mmol (Amersham Biosciences Corporation, Piscataway, NJ). The housekeeping control, EF-1, was used to evaluate the spot-to-spot variability within the experiment. See Farkas *et al.*, 2003, *J. Biol. Chem.* 384: 945 for grid wherein the positional coordinates of the array are defined.

### 6.2.9. CANCER PROFILING ARRAY (CPA™)

**[0396]** The CPA™ (Clontech, Palo Alto, CA) was used to determine the expression of SGA-56M and SGA-56Mv in numerous tumor/normal paired patient samples. The CPA™ contains 241 tumor and adjacent normal paired patient isolates. Primers used in amplifying a probe were common for SGA-56M and SGA-56Mv. Primers were as follows: RT1 (5'-GCTTGGAAAAGTTGAGCC -3') (SEQ ID NO: 19), and RT2 (5'-

CTGGGTCTGAGTCTTAGC-3') (SEQ ID NO; 20). Fifty ng of PCR product was labeled using Ready-to-go Beads and α-P32 dCTP at 3000 Ci/mmol. A total of 241 paired cDNA samples were synthesized and spotted onto nylon membranes for 13 different tumor types. The tumor types included: Breast, Cervix, Colon, Kidney, Lung, Ovarian, Pancreas, Prostate, Rectum, Thyroid Gland, Small Intestine, Stomach, and Uterus. See Zhumabayeva *et al.*, 2000, *BioTechniques*. 3: 22 for grid wherein the positional coordinates of the array are defined.

### 6.2.10 ABI PRISM®7000 SEQUENCE DETECTION SYSTEM

**[0397]** The ABI PRISM® 7000 Real-Time PCR Sequence Detection System (Applied Biosystems, Foster City, CA) was used to determine the breast cancer-selectivity for SGA-56M/SGA-56Mv. The Breast Cancer Rapid-Scan™ gene expression RNA panel (OriGene Technologies, Inc., Rockville, MD) and Lung Cancer patient tumor and adjacent normal tissue RNA (Biochain Institute, Inc., Hayward, CA and Ambion, Inc., Austin, TX) were used in this experiment. The Rapid-Scan™ Panel contains first-strand cDNA derived from 12 tumor/normal breast patient sample pairs. Lung Cancer cDNA was synthesized from using the Superscript First-Strand cDNA Synthesis System for RT-PCR® (Invitrogen, Carlsbad, CA). Primers and probes for SGA-56M and SGA-56Mv were as follows: EXP2-FP (5-TGTCCCAGGAACCTTTCTTCA-3') (SEQ ID NO: 21), EXP2-RP (5'-CCCAGCTTGCACCTGGTTT-3') (SEQ ID NO: 22), and EXP2-TaqMan MGB Probe (5'-FAM-CTACAGCTCACTCTCCAG-NFQMGB-3') (SEQ ID NO: 23). Primers and probes for EF1 were as follows: EF1-FP (5'-ATGACCCACCAATGGAAGCA-3') (SEQ ID NO: 24), EF1-RP (5'-GCCTGGATGGTTCAGGATAATC-3') (SEQ ID NO: 25), and EF1-TaqMan MGB Probe (5'-VIC-CTGGCTTCACTGCTC-NFQMGB-3') (SEQ ID NO: 26). EF-1 was used as the normalization gene for all ABI PRISM® 7000 experiments.

**[0398]** The Comparative Ct Method (Applied Biosystems, Foster City, CA) was used in calculating tumor vs. normal ratios for SGA-56M/SGA-56Mv. The amount of target (SGA-56M/SGA-56Mv), normalized to an endogenous reference (EF-1) and relative to a calibrator, is given by the arithmetic formula: $2^{-\Delta\Delta Ct}$ where ΔΔCt is the change in threshold cycle between target and reference.

### 6.2.11. BIOINFORMATICS ANALYSIS

**[0399]** After completion of the array data analysis sorting process, interesting novel targets were retained and analyzed using several additional computational programs. The derived SGA-56M and SGA-56Mv cDNA was analyzed using Vector NTI Suite 6.0® (InforMax, Inc., Bethesda, MD). Transmembrane domain and protein localization analysis were performed using the ExPASy Proteomics Tools Programs® (Swiss Institute of Bioinformatics, Geneve, Switzerland). Amino acid sequence prediction programs used included: HMMTOP (Tusnady *et al.*, 1998, *J. Mol. Bio.* 283:489), TM pred (Hofinann *et al.*, 1993, *J. Biol. Chem.* 347:166), TMHMM v1.0 (Sonnhammer *et al.*, 1998, *Proc. of Sixth Int. Conf. on Intelligent Systems for Mol. Bio.*, AAAI Press, pp. 175-182), TMAP, and PSORT (Nakai *et al.*, 1999, *Trends Biochem. Sci.* 24(1):34).

### 6.2.12. SUBCELLULAR LOCALIZATION OF SGA-56M AND SGA-56Mv

**[0400]** The subcellular localization patterns for SGA-56M and SGA-56Mv were determined using green fluorescent protein (GFP) reporter constructs. SGA-56M and SGA-56Mv cDNA clones were amplified by PCR using gene-specific primers: SGA-56M (GFP1) 5'-AGCTCTCTCGAGATGTCTTTTCTTGGCATCCTGTGCAAGTGT-3' (SEQ ID NO: 27) and SGA-56M (GFP2) 5'- AGCTCTAAGCTTTCAGTGTGGAGGGTTCATGGTGCCTTG-3' (SEQ ID NO: 28). *Xho* I (SGA-56M-GFP1) and *Hind* III (SGA-56M-GFP2) restriction sites were included for in-frame cloning (as underlined). The resulting SGA-56M and SGA-56Mv PCR products were restriction digested and cloned into the *Xho* I/*Hind* III-cut pGFP$^2$ vector (BioSignal Packard, Montreal, Canada). Expression of this plasmid in eukaryotic cells resulted in the synthesis of both SGA-56M/GFP and SGA-56Mv/GFP fusion polypeptides. These constructs were transiently transfected into human kidney 293 cells, and SKBR-3 breast carcinoma cells by electroporation. The subcellular localization patterns for SGA-56M and SGA-56Mv green fluorescence signals were monitored by fluorescence microscopy.

## 6.3 RESULTS

### 6.3.1. ISOLATION OF THE SGA-56M cDNA

**[0401]** The SGA-56M cDNA (FIG. 1) was amplified using gene-specific primers and cloned into the pCR 4.0® TOPO TA vector (Invitrogen, Carlsbad, CA). The SGA-56M sequence (FIG. 1) (SEQ ID NO: 1) was sequence verified using custom primers (Sigma-Genosys, Woodlands, TX) and automated fluorescent sequencing (PE Applied Biosystems, Foster City, CA).

### 6.3.2. ISOLATION OF THE SGA-56Mv cDNA

**[0402]** The SGA-56Mv cDNA (FIG. 2) was amplified using gene-specific primers and cloned into the pCR 4.0® TOPO TA vector (Invitrogen, Carlsbad, CA). SGA-56Mv (SEQ ID NO: 3) was identified while screening clones for the isolation of SGA-56M (SEQ ID NO: 1). SGA-56Mv (FIG. 2) was sequence verified using custom primers (Sigma-Genosys, Woodlands, TX) and automated fluorescent sequencing (PE Applied Biosystems, Foster City, CA).

### 6.3.3. CANCER-SELECTIVITY BY SEMI-QUANTITATIVE PCR

**[0403]** SGA-56M and SGA-56Mv displayed cancer-selectivity on various breast carcinoma cell-lines (FIG. 3). A cDNA region common to both SGA-56M and SGA-56Mv was amplified in this experiment. All breast cancer cell-lines evaluated were positive for SGA-56M and SGA-56Mv mRNA (FIG. 3). Normal human mammary epithelial cells (HMECs) were negative for SGA-56M and SGA-56Mv mRNA expression even after 35 PCR cycles.

**[0404]** SGA-56M and SGA-56Mv displayed positive mRNA expression in other tumor cell-lines (FIG. 4). Positive tumor cell-lines for SGA-56M and SGA-56Mv mRNA include: Ramos (Burkitt's lymphoma), NCI-

H460 (Non-Small Cell Lung Cancer, NSCLC), MiaPaCa-2 (Pancreatic Cancer), and WM-115 (Melanoma) (FIG. 4).

### 6.3.4. EVALUATION OF NORMAL EXPRESSION BY MTE™

**[0405]** SGA-56M and SGA-56Mv mRNA expression levels in normal tissues were evaluated using the Multiple Tissue Expression (MTE™) Array. A cDNA region common to both SGA-56M and SGA-56Mv was amplified and used as a probe for this experiment. The MTE™ Array contains 76 tissue-specific polyA+ RNA isolates. See Farkas *et al.*, 2003, *J. Biol. Chem.* 384: 945 for a grid of the tissue represented on the array. SGA-56M and SGA-56Mv displayed minimal expression on normal tissue. The only significant level of normal tissue expression was observed in the testis, at position 8F (FIG. 5B).

### 6.3.5. CANCER-SELECTIVITY BY CPA™

**[0406]** SGA-56M and SGA-56Mv mRNA cancer-selectivity was evaluated using the Cancer Profiling Array (CPA™). A cDNA region common to both SGA-56M and SGA-56Mv was amplified and used as a probe for this experiment. SGA-56M and SGA-56Mv displayed cancer-selective expression greater than 2-fold in several tumor types (Table 3). Breast tumor and corresponding normal tissue pairs displaying the highest T: N ratios for SGA-56M are illustrated (Table 4). SGA-56M and SGA-56Mv probes were also cancer-selective in eight patient isolates known to have associated metastases.

Table 3

*SGA-56M and SGA-56Mv cancer-selectivity in individual tumor and corresponding normal tissues*

| A. Tumor Tissues | >2-fold T: N |
|---|---|
| Breast (n = 50) | 30% |
| Uterus (n = 42) | 19% |
| Colon (n = 35) | 14% |
| Stomach (n = 27) | 7% |
| Ovary (n = 14) | 29% |
| Lung (n = 21) | 14% |
| Kidney (n = 20) | 5% |
| Rectum (n = 18) | 33% |

| B. Normal Tissues | >2-fold N: T |
|---|---|
| Breast (n = 50) | 0% |
| Uterus (n = 42) | 2% |
| Colon (n = 35) | 0% |
| Stomach (n = 27) | 0% |
| Ovary (n = 14) | 7% |
| Lung (n = 21) | 0% |
| Kidney (n = 20) | 0% |
| Rectum (n = 18) | 0% |

Table 4

*Elevated breast cancer-selectivity of SGA-56M and SGA-56Mv in a subset of patients*

| Tumor type | Age | T: N |
|---|---|---|
| Infiltrating ductal carcinoma | 52 | 9 |
| Infiltrating ductal carcinoma | 45 | 5 |
| Infiltrating ductal carcinoma | 44 | 5 |
| Infiltrating ductal carcinoma | 60 | 4 |
| Infiltrating lobular carcinoma | 49 | 4 |
| Infiltrating lobular carcinoma | 66 | 4 |
| Medullary carcinoma | 47 | 3 |
| Adenocarcinoma | 53 | 3 |
| Tubular carcinoma | 63 | 3 |
| Fibrosarcoma | 44 | 3 |

### 6.3.6. CANCER-SELECTIVITY BY ABI PRISM® 7000 REAL-TIME PCR

**[0407]** SGA-56M and SGA-56Mv displayed breast cancer-selectivity using the breast cancer Rapid Scan™ cDNA panel (Table 6). Real-time PCR was used to further quantify the extent of over-expression of SGA-56M and SGA-56Mv in breast and lung patient tumor isolates. Twelve breast tumor and corresponding normal tissues were analyzed for quantitative SGA-56M and SGA-56Mv expression levels (Table 5). The comparative $C_T$ method was used in calculating relative quantitative T: N ratios while using the control gene EF-1 as a reference. In total, 5 of 12 breast cancer patient pairs (42%) displayed T: N levels > 3-fold (Table 6). A single cDNA pair (sample #10) displayed a T: N ratio of 14.1 (Table 6).

**[0408]** SGA-56M and SGA-56Mv lung cancer-selectivity was examined using 10 non small-cell lung cancer patient pairs, 5 adenocarcinomas and 5 squamous cell carcinomas with corresponding normal tissues (Table 7). Quantitative T: N ratios were calculated using methods as described above. In total, 7 of 10 non small-cell lung cancer patients (70%) displayed T: N levels > 3-fold (Table 8). In particular, all of the squamous cell carcinoma patients appeared to exhibit elevated expression levels > 4-fold T: N (Table 8). Two cDNA pairs (SQ1 and SQ4) displayed T: N levels > 10-fold (Table 8).

Table 5

*Histopathology data for breast cancer QPCR tissue panel*

| Sample | Tumor Type | ER/PR | ER (fmol/mg) | PR (fmol/mg) |
|---|---|---|---|---|
| 1 | Invasive mixed tubular carcinoma | ER+/PR+++ | 7 | 233 |
| 2 | Invasive ductal carcinoma | ER+/PR+++ | 14 | 99 |
| 3 | Invasive lobular carcinoma | ER+++++/PR+++++ | 142 | 528 |
| 4 | Invasive ductal carcinoma | ER++/PR- | 20 | 9 |
| 5 | Invasive ductal carcinoma | ER++/PR- | 18 | 7 |
| 6 | Invasive ductal carcinoma | ER+++/PR+ | 65 | 30 |
| 7 | Invasive ductal carcinoma | ER++/PR+ | 30 | 32 |
| 8 | Invasive ductal carcinoma | ER+/PR- | 9 | 0.5 |
| 9 | Adenoid cystic carcinoma | ER++/PR+ | 22 | 14 |
| 10 | Invasive ductal carcinoma | ER-/PR- | 3 | 0 |
| 11 | Ductal carcinoma in-situ | ER+/PR+ | 19 | 13 |
| 12 | Invasive ductal carcinoma | ER+/PR+ | 6 | 26 |

Table 6

*SGA-56M and SGA-56Mv breast cancer-selectivity in patient tumors by QPCR*

| Sample | SGA-56M Ct | EF-1 Ct | ΔCt | ΔΔCt | T: N |
|---|---|---|---|---|---|
| Breast Tumor 1 (ER+) | 31.13 | 30.35 | 0.78 | -0.97 | 1.96 |
| Breast Normal 1 | 30.42 | 28.67 | 1.75 | | |
| Breast Tumor 2 (ER+) | 31.05 | 29.61 | 1.44 | -0.58 | 1.49 |
| Breast Normal 2 | 30.83 | 28.81 | 2.02 | | |
| Breast Tumor 3 (ER+) | 30.40 | 29.51 | 0.89 | -2.23 | 4.70 |
| Breast Normal 3 | 31.69 | 28.57 | 3.12 | | |
| Breast Tumor 4 (ER+) | 31.18 | 29.40 | 1.78 | -0.57 | 1.48 |
| Breast Normal 4 | 31.82 | 29.47 | 2.35 | | |
| Breast Tumor 5 (ER+) | 30.55 | 27.61 | 2.94 | 1.46 | 0.36 |
| Breast Normal 5 | 30.79 | 29.31 | 1.48 | | |
| Breast Tumor 6 (ER+) | 31.42 | 30.82 | 0.60 | -2.21 | 4.63 |
| Breast Normal 6 | 31.08 | 28.27 | 2.81 | | |
| Breast Tumor 7 (ER+) | 31.57 | 30.86 | 0.71 | -1.46 | 2.75 |
| Breast Normal 7 | 31.78 | 29.61 | 2.17 | | |
| Breast Tumor 8 (ER+) | 30.80 | 30.94 | -0.14 | -1.75 | 3.36 |
| Breast Normal 8 | 30.41 | 28.80 | 1.61 | | |
| Breast Tumor 9 (ER+) | 31.26 | 29.47 | 1.79 | -0.14 | 1.10 |
| Breast Normal 9 | 29.54 | 27.61 | 1.93 | | |
| Breast Tumor 10 (ER-) | 30.94 | 32.33 | -1.39 | -3.82 | 14.12 |
| Breast Normal 10 | 30.82 | 28.39 | 2.43 | | |
| Breast Tumor 11 (ER+) | 30.92 | 28.71 | 2.21 | 1.09 | 0.47 |
| Breast Normal 11 | 30.07 | 28.95 | 1.12 | | |
| Breast Tumor 12 (ER+) | 30.30 | 27.90 | 2.40 | 0.81 | 0.57 |
| Breast Normal 12 | 30.61 | 29.02 | 1.59 | | |

Table 7

*Background information for lung cancer QPCR tissue panel*

| Sample | Tumor Type | Differentiation | Age | Sex |
|---|---|---|---|---|
| Lung Tumor AD01 | Adenocarcinoma | Moderately Differentiated | 44 | M |
| Lung Tumor AD02 | Adenocarcinoma | Poorly Differentiated | 62 | M |
| Lung Tumor AD03 | Adenocarcinoma | Poorly Differentiated | 58 | F |
| Lung Tumor AD04 | Adenocarcinoma | Moderately Differentiated | 60 | M |
| Lung Tumor AD05 | Adenocarcinoma | Moderately Differentiated | 73 | F |
| Lung Tumor SQ01 | Squamous Cell Carcinoma | Well Differentiated | 78 | M |
| Lung Tumor SQ02 | Squamous Cell Carcinoma | Well Differentiated | 62 | M |
| Lung Tumor SQ03 | Squamous Cell Carcinoma | Moderately Differentiated | 63 | F |
| Lung Tumor SQ04 | Squamous Cell Carcinoma | Poorly Differentiated | 64 | M |
| Lung Tumor SQ05 | Squamous Cell Carcinoma | Moderately Differentiated | 66 | M |

Table 8

*SGA-56M and SGA-56Mv lung cancer-selectivity in patient tumors by QPCR*

| Sample | SGA-56M Ct | EF-1 Ct | ΔCt | ΔΔCt | T: N |
|---|---|---|---|---|---|
| Lung Tumor AD01 | 33.43 | 25.91 | 7.52 | -0.45 | 1.36 |
| Lung Normal | 34.08 | 26.11 | 7.97 | | |
| Lung Tumor AD02 | 33.78 | 26.81 | 6.97 | -2.70 | 6.50 |
| Lung Normal | 35.72 | 26.05 | 9.67 | | |
| Lung Tumor AD03 | 36.43 | 24.94 | 11.49 | 2.00 | 0.25 |
| Lung Normal | 34.62 | 25.13 | 9.50 | | |
| Lung Tumor AD04 | 36.05 | 25.41 | 10.64 | 1.30 | 0.41 |
| Lung Normal | 35.48 | 26.14 | 9.34 | | |
| Lung Tumor AD05 | 35.29 | 30.24 | 5.05 | -1.74 | 3.34 |
| Lung Normal | 36.48 | 29.69 | 6.79 | | |
| Lung Tumor SQ01 | 33.70 | 27.60 | 6.10 | -3.48 | 11.16 |
| Lung Normal | 36.10 | 26.52 | 9.58 | | |
| Lung Tumor SQ02 | 35.24 | 30.12 | 5.12 | -2.72 | 6.57 |
| Lung Normal | 37.04 | 29.21 | 7.83 | | |
| Lung Tumor SQ03 | 32.96 | 28.35 | 4.61 | -2.84 | 7.16 |
| Lung Normal | 34.84 | 27.39 | 7.45 | | |
| Lung Tumor SQ04 | 34.43 | 27.41 | 7.02 | -3.73 | 13.22 |
| Lung Normal | 36.92 | 26.18 | 10.74 | | |
| Lung Tumor SQ05 | 37.97 | 32.07 | 5.90 | -2.03 | 4.07 |
| Lung Normal | 39.25 | 31.33 | 7.93 | | |

### 6.3.7. SEQUENCE COMPARISON FOR SGA-56M AND SGA-56Mv

**[0409]** Nucleic acid entries sharing homology with SGA-56M include: GenBank Accession No. D87437, GenBank Accession No. NM_014837, GenBank Accession No. AB085674, GenBank Accession No. AX714019, and GenBank Accession No. AX747010.

**[0410]** GenBank Accession No. D87437, and GenBank Accession No. NM_014837 are termed KIAA0250 in the scientific literature (Nagase *et al.*, 1996, *DNA Research.* 3(5): 321-329, 341-354. and Sood *et al.*, 2001, *Genomics*, 73:211-222). KIAA0250 does not share any significant sequence homology to known motifs that would suggest a potential functional role (Sood *et al.*, 2001, *Genomics*, 73:211-222). KIAA0250 was identified as 1 of 13 novel transcripts that mapped to the hereditary prostate locus (HPC1) (Sood *et al.*, 2001, *Genomics*, 73:211-222).

**[0411]** GenBank Accession No. AB085674, termed Homo sapiens mRNA SMG-7, also shares homology with SGA-56M and SGA-56Mv. To date, no putative biological role for SMG-7 has been described in the scientific literature. SMG-1, a novel member of the phosphatidylinositol 3-kinase family of proteins, has been reported to be associated with nonsense-mediated mRNA decay (NMD), due presumably to its ability to phosphorylate hUpf1 (Denning *et al.*, 2001, *J. Biol. Chem.* 276(25): 22709-22714). Neither SMG-1 nor SMG-7 has been described as associated with cancer.

[0412] GenBank Accession No. AX714019 and GenBank Accession No. AX747010 correspond to SEQ ID 703 (European Patent Application EP1293569) and SEQ ID 535 (European Patent Application EP1308459), respectively. These applications disclose molecules (AX714019 and AX747010) that appear to correspond to partial cDNA sequences having limited homology to full length SGA-56M (SEQ ID NO: 1) and SGA-56Mv (SEQ ID NO: 3). See Table 9. As indicated herein above, SEQ ID NO: 1 and SEQ ID NO: 3 have been deposited in GenBank.

Table 9

| | | Alignment | |
|---|---|---|---|
| Gene | Nucleic Acids | AX714019 (%) | AX747010 (%) |
| SGA-56M | 2917 | 1755/2917 (60%) | 1753/2917 (60%) |
| SGA-56Mv | 2779 | 1617/2779 (58%) | 1615/2779 (58%) |

### 6.3.8 SUBCELLULAR LOCALIZATION OF SGA-56M AND SGA-56Mv

[0413] Subcellular localization patterns for SGA-56M and SGA-56Mv were determined using fluorescence microscopy. Transient expression and subcellular localization pattern recognition of SGA-56M/GFP and SGA-56Mv/GFP constructs were analyzed using 293 human kidney cells, and SKBR-3 breast carcinoma cells. Expression of GFP alone resulted in diffuse green fluorescence signals throughout the cells. Subcellular localization patterns for SGA-56M/GFP and SGA-56Mv/GFP were consistent with those previously reported for other proteins of cytoplasmic and/or peroxisomal compartments (Simpson *et al.*, 2000, *EMBO reports*, 3: 287-292).

## 6.4. DISCUSSION

[0414] Gene expression profiling provides a systematic approach to studying the mechanisms associated with progression from normal to metastatic disease. In this application, the present inventors have combined SSH and cDNA microarrays to identify the uncharacterized breast cancer-associated antigen, SGA-56M and variants thereof, including SGA-56Mv. Combining SSH and cDNA microarrays provides a rapid and effective approach to high-throughput screening for novel tumor associated antigens (TAAs). The principle of SSH allows for the preferential amplification of differentially expressed sequences while suppressing those present at equal abundance within the initial mRNA (Diatchenko *et al.*, *supra*). The high level of enrichment, low level of background, and efficient normalization of sequences makes this an attractive approach for the rapid identification of novel targets. SGA-56M cDNA, identified by this method, and variants thereof, including SGA-56Mv comprise new diagnostic, prognostic, and/or therapeutic targets for breast and lung cancer treatment. SGA-56M and SGA-56Mv display tumor-selective expression in breast and lung cancer, and other cancers, while displaying minimal expression in normal tissues. SGA-56M and SGA-56Mv, based on their elevated level of tumor-selective expression, and association with metastases are strong candidates for consideration and evaluation as a potential mode of intervention for breast cancer and other cancers.

[0415] Overall, SGA-56M and SGA-56Mv, can be helpful in providing valuable insight into the potential mechanisms involved in breast cancer development and progression. The present inventors have demonstrated that gene expression profiling studies using SSH and arrays can assist in identifying interesting cancer-selective genes, like SGA-56M and SGA-56Mv, that have not been previously implicated in breast cancer. Studies to investigate further the potential functional role of tumor associated antigens are extremely helpful in designing experiments to critically evaluate the pathways and mechanisms necessary for effective therapeutic intervention.

## 7. REFERENCES CITED

[0416] All references cited herein are incorporated herein by reference in their entirety and for all purposes to the same extent as if each individual publication or patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety for all purposes.

[0417] Many modifications and variations of this invention can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. The specific embodiments described herein are offered by way of example only, and the invention is to be limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled.